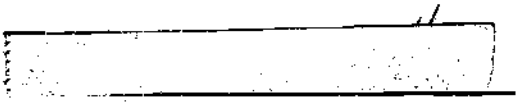
In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.



7/25/68

DEVELOPMENT OF THE MULTIPLEXED GRADIENT TECHNIQUE WITH APPLICATIONS TO THE EXTREMUM CONTROL OF MULTISTAGE CHEMICAL REACTORS

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

Richard Fox Barrett

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

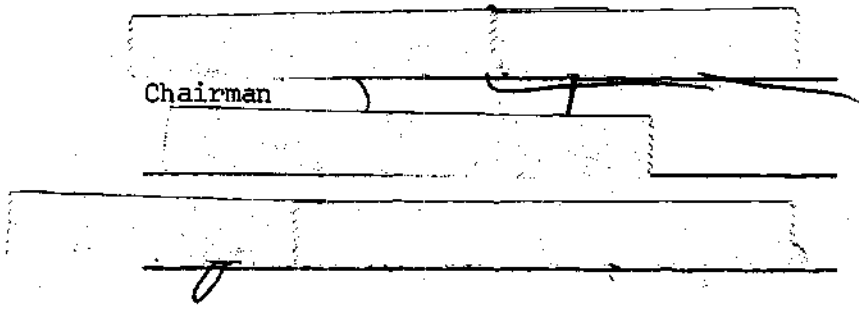
in the School of Electrical Engineering

Georgia Institute of Technology

June, 1971

DEVELOPMENT OF THE MULTIPLEXED GRADIENT TECHNIQUE WITH APPLICATIONS TO THE EXTREMUM CONTROL OF MULTISTAGE CHEMICAL REACTORS

Approved:



Chairman

Date approved by Chairman: May 25, 1971

ACKNOWLEDGMENTS

Grateful acknowledgment is extended to Dr. Cecil O. Alford, my advisor. The time, effort, and insight he has afforded me in this dissertation are deeply appreciated. Appreciation is also extended to Dr. Charles W. Gorton for his many helpful suggestions during the course of the research. Gratitude is likewise extended to Dr. James R. Rowland for his participation as a member of the reading committee.

Sincere appreciation is extended to Dr. Joseph L. Hammond, Jr., for originally suggesting the thesis problem.

Appreciation is extended to the Lockheed-Georgia Company for providing hybrid computing time during the early stages of the research, and to the School of Electrical Engineering for providing digital computing time throughout the doctoral research. Special thanks are extended to Dr. George Burnet of the Iowa State University of Science and Technology for lending his personal copy of a much needed reference. Finally, appreciation is extended to the Department of Health, Education, and Welfare for the NDEA Fellowship which made possible the doctoral studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

The chemical industry has long used the skill of the human operator to adjust process set points to optimize performance, but only in the last few years have small, reliable digital computers been available for this task. Central to the use of the digital computer is a hierarchal control structure, with processes regulated by set point controllers, set points established by operational control algorithms, and scheduling of start-up and shut-down dictated by supervisory control algorithms. One approach to process optimization, implemented as a part of the operational control software, is to adjust controller set points to maximize a given measure of process performance.

This dissertation outlines the development of such a control algorithm for a class of nonlinear time-varying distributed-parameter processes with multiple inputs and multiple outputs. The class of systems considered is described by the vector partial differential equation

$$\frac{\partial \underline{x}(\ell,t)}{\partial t} + v(\ell,t)\,\frac{\partial \underline{x}(\ell,t)}{\partial \ell} = \underline{f}[\underline{x}(\ell,t),u(\ell,t),\ell,t]$$

where $\underline{x}(\ell,t)$ is the vector state of a fluid flowing through the system with a velocity $v(\ell,t)$, $u(\ell,t)$ is the system control, ℓ is the spatial variable defined on $0<\ell\leq L$, and t is time defined on $t\geq 0$. This equation may be used to describe the behavior of multijacketed tubular reactors

and, in a degenerate form, multistage stirred-tank reactor systems. The systems considered are assumed to be controlled by set point controllers and to be constrained to satisfy magnitude constraints on the controller set points and on certain monitored system variables. Constraints are also imposed on the repetition rate of the system measuring instrument. The control objective is to adjust the set points, subject to these constraints, to maximize a given function of monitored system states.

The algorithm developed is a fixed-increment stepwise extremum control procedure. It differs from existing procedures in its use of system transport delays to interlace the effects of several set point combinations. Hence the name Multiplexed Gradient Technique. More specifically, the algorithm carefully coordinates the times at which set points are changed and measurements are made so that a separation is maintained between fluid elements representative of different system states. These states are then sampled at appropriate times, and the data obtained used to generate a steady-state locally-linear system model. From this model, new set points are selected which increase the value of the function to be maximized. The cycle of model generation and set point adjustment is repeated continually, with set points first achieving and then tracking a drifting system optimum.

Presented in the dissertation is the development of the basic control algorithm and the extension of that algorithm to the control-constrained and state-constrained cases. The effects of measurement noise are examined, and the hardware required to implement the algorithm is established. Finally, the performance of the algorithm is compared

to that of two currently available procedures, using simulated tubular and stirred-tank reactor systems as examples. Through these studies the algorithm is shown to have excellent speed, convergence, tracking, and noise insensitivity properties.

The major contributions of the research are the unification of a number of diverse system types and control objectives into a common mathematical framework, and the development of a control algorithm which uses to advantage the transport delays inherent in these systems.

CHAPTER I

INTRODUCTION

The chemical industry has long used the skill of the human operator to adjust process set points to optimize performance, but only in the last few years have small, reliable digital computers been available for this task. Central to the use of the digital computer has been a hierarchal control structure, with processes regulated by set point controllers, set points established by operational control algorithms, and scheduling of start-up and shut-down dictated by supervisory control algorithms.[1,2] One approach to process optimization, implemented as a part of the operational control software, has been to adjust control set points to maximize a given measure of process performance.

The objective of the thesis research was to develop such an extremum control algorithm for a class of nonlinear time-varying distributed-parameter processes with multiple inputs and multiple outputs. The class of systems considered included several types of multistage chemical reactors used extensively in the manufacture of organic chemicals. Given below is a review of the characteristics of these systems, followed by an examination of techniques available for their optimization.

Multistage Chemical Reactors

Consider as an example the production of ethylene oxide from ethylene and oxygen. Stoichiometrically, the reaction may be described by[3,4]

$$2\ C_2H_4 + O_2 \longrightarrow 2\ C_2H_4O. \tag{1-1}$$

This reaction may be accompanied by the undesired side reaction[3,4]

$$C_2H_4 + 3\ O_2 \longrightarrow 2\ CO_2 + 2\ H_2O \tag{1-2}$$

producing carbon dioxide and water. Since the rates at which these two reactions take place are different functions of temperature, proper temperature control can suppress the side reaction and favor the production of ethylene oxide. Thus, if $x_1(t)$ is the concentration of C_2H_4O, $x_2(t)$ the concentration of CO_2, and $u(t)$ the temperature of the mixture, the progress of the reaction may be described by a vector ordinary differential equation of the form

$$\frac{d\underline{x}(t)}{dt} = \underline{f}[\underline{x}(t),u(t)] \qquad 0 \le t \le T \tag{1-3}$$

$$\underline{x}(0) = \underline{x}_o.$$

A straightforward computation in the calculus of variations, assuming that all of the parameters of (1-3) are known, establishes the

temperature profile $u^*(t)$, $0 \le t \le T$, which maximizes $x_1(T)$, the final concentration of ethylene oxide.

To achieve production on a continuous rather than a batch basis, the mixture may be moved through a spatially-distributed vessel with a velocity v. Such a vessel is shown in Figure 1. With the transformation $t = \ell/v$, where ℓ is the spatial dimension of the vessel and v is assumed constant, (1-3) describes the steady-state conditions along the length of the vessel. Optimal conditions are achieved by maintaining a temperature given by $u^*(\ell/v)$, $0 \le \ell \le L$. Rarely can this spatial temperature profile be achieved, for the control required is prohibitively expensive. Multistage chemical reactors provide an economically feasible means of approximating $u^*(\ell/v)$, either by segmenting $u(\ell/v)$ into several piecewise-constant elements, as in the multijacketed tubular reactor, or by segmenting the entire system so that the governing equations become difference equations, as in the sequence of stirred-tank reactors. Schematic diagrams of these multistage systems are given in Appendix A. Typical industrial systems involve two or three stages, with the time required for the reactants to move through all stages on the order of 30 seconds for gaseous reactions and 30 minutes for liquid reactions.

Techniques for Reactor Optimization

Research into the computation of the profiles $u^*(\ell/v)$ and empirical investigations into the optimization of multistage systems began in the 1950's.[5-7] Only in the last decade, however, have techniques been available for computing optimal steady-state conditions in wide classes

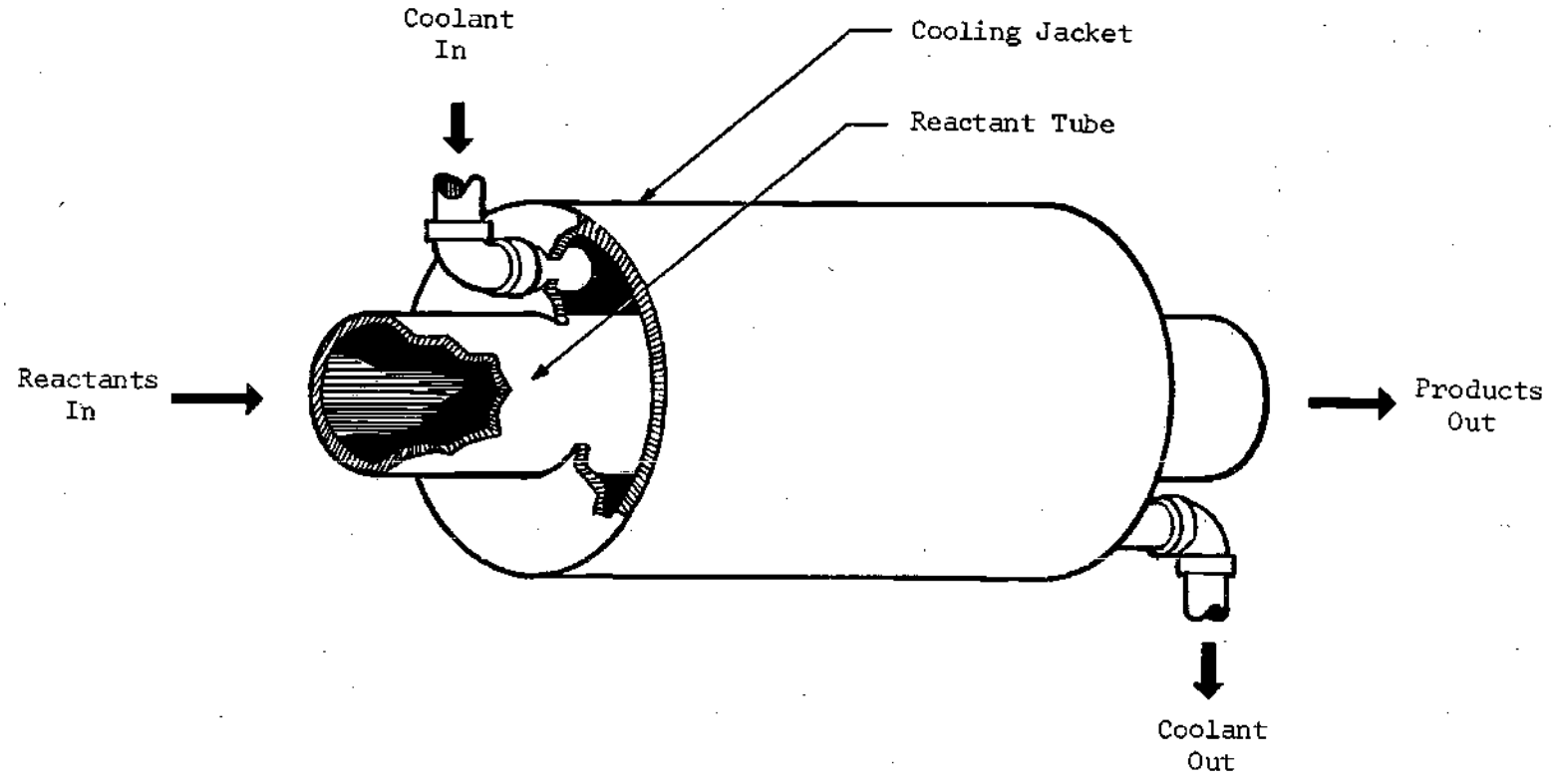


Figure 1. Tubular Reactor with Cooling Jacket

of multistage systems, such as tubular reactors with multiple cooling jackets,[8] adiabatic catalytic reactors with heat exchangers between stages,[9] sequences of stirred-tank reactors,[10] and the series combination of tubular and stirred-tank reactors.[11]

The mathematical models used in these computations have not been exact, due to unknown and time-varying system parameters, and consequently the set points computed have been only approximately optimal. One approach to the problem of changing system parameters has been to use model-reference adaptive controllers to adjust the set points as the parameters drifted, thereby minimizing the deviation of the system from the model.[12] Another has been to track the drifting optimum by computing time-varying optimal set points from a time-varying model.[13,14] Optimal control theory has been used to design feedforward controllers,[15] and state estimation theory has been used to estimate system parameters.[16]

The most widely-accepted approach to the problem of unknown and time-varying parameters has been to continually perturb the set points, continually monitor the function to be maximized, and determine from the measured data the required set point changes. Rather simple approaches have been used on systems with only one controlled variable.[17,18] For more complex systems, there have been two basic approaches. The first has been to apply a sinusoidal perturbation of a different frequency to each set point.[19-21] The resulting fluctuation in the function to be maximized has then been filtered with bandpass filters, and the magnitude and phase of each filter output, relative to that of the

corresponding set point perturbation, used to determine the required set point change. The second approach has been to apply a step perturbation to one or more set points, wait for the system to reach a new steady state, and then measure the function to be maximized.[22] Each such measurement has defined a point in (n+1)-dimensional Euclidean space, where n was the number of set points. By extrapolating a surface passed through several such points, new set points have been selected which increased the value of the function to be maximized. The chemical industry has favored this technique over the sinusoidal perturbation method, for with this technique the time lags present in so many chemical systems have not created instability in the loop containing the optimizing controller.

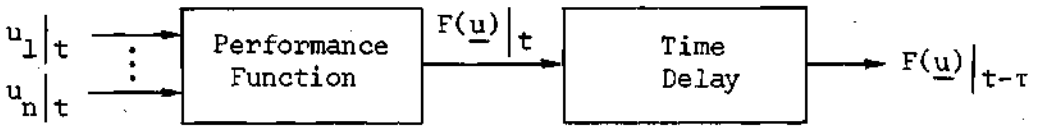


Figure 2. System Model for Currently Available Algorithms

Currently available stepwise extremum control algorithms have been developed from the system model shown in Figure 2.[23,24] The first element of this model is a steady-state performance function which maps the n set points $\underline{u} = \{u_1,\ldots,u_n\}$ into the function $F(\underline{u})$ to be maximized. The second is a time delay τ which represents the time required for the system to reach steady state after a change in set points. Following any change in set points, currently available algorithms have measured $F(\underline{u})$ only after a time τ has elapsed.

The approach used by most techniques has been to assume that $F(\underline{u})$ may be represented by the linear model

$$F(\underline{u}) = F(\underline{u}^0) + \sum_{j=1}^{n} \left.\frac{\partial F(\underline{u})}{\partial u_j}\right|_{\underline{u}^0} (u_j - u_j^0) \tag{1-4}$$

in a neighborhood of an initial point $\underline{u}^0$. Measured system data has then been used to approximate the partial derivatives of (1-4). One approximation which has been used is

$$\left.\frac{\partial F(\underline{u})}{\partial u_j}\right|_{\underline{u}^0} = \frac{1}{\Delta u_j} [F(u_1^0,\ldots,u_{j-1}^0,u_j^0+\Delta u_j,u_{j+1}^0,\ldots,u_n^0) - F(\underline{u}^0)]. \tag{1-5}$$

Various techniques have been developed for choosing an adjustment vector $\Delta\underline{u}$ so that the new set points

$$\underline{u}^1 + \underline{u}^0 + \Delta\underline{u} \tag{1-6}$$

result in an increase in $F(\underline{u})$:

$$F(\underline{u}^1) > F(\underline{u}^0). \tag{1-7}$$

The standard gradient technique, for example, invests n measurements into the approximations of (1-5), and then chooses the components of $\Delta\underline{u}$ proportional to the respective $\partial F(\underline{u})/\partial u_j$.[25] Other techniques adjust

one set point at a time[26] or fix the magnitude of the increments in the set points.[27]

As the set points approach the optimum, the $\partial F(\underline{u})/\partial u_j$ become quite small and measurement noise prevents their accurate determination. This noise, coupled with a breakdown in the linear model (1-4), results in rather erratic limit cycles about the optimum.[28] Since the $\partial F(\underline{u})/\partial u_j$ are small, these limit cycles cause no significant degradation in $F(\underline{u})$ and are generally allowed to continue. Extremum control algorithms designed to filter the noise effects have been considered,[29] but massive computational requirements have discouraged their use.

When constraints are present the procedures outlined above must be modified, for the calculated $\underline{u}^1$ may violate an explicit constraint on the set points or may cause some system variable $G(\underline{u})$ to exceed a preset limit. The most common procedural modification has been to terminate the adjustment of any set point whose change threatened to cause a constraint violation. Another has been to force selection of set points away from the constraint boundaries by penalizing $F(\underline{u})$ whenever $G(\underline{u})$ approached a boundary.[18] A procedure has been proposed which paired one set point with each $G(\underline{u})$ encountering a constraint boundary.[30] This set point was then adjusted by a regulatory control algorithm, rather than an extremum control algorithm, to force $G(\underline{u})$ to track the constraint boundary. Unfortunately, the method did not always converge to the optimum. The techniques of nonlinear programming[31,32] do provide the necessary convergence properties, and their use with one of the more robust extremum control algorithms has recently been suggested.[27]

In summary, considerable effort has been devoted to the computation of optimal steady-state conditions in multistage chemical reactors, and various techniques have been used to combat the problem of unknown and time-varying system parameters. None of these techniques, however, has had any means of handling system transport delays except to treat the system as a lumped-parameter process with an excessively long transient settling time.

Synopsis of the Research

The objective of the thesis research was to develop an extremum control algorithm which would adequately handle system transport delays. To this end, a system model was developed which was considerably more detailed than that given in Figure 2 (page 6). A stepwise extremum control algorithm was then developed which utilized system transport delays to interlace the effects of several set point combinations. More specifically, the new system model was used to coordinate the times at which set points were changed with the times at which system measurements were made. In this manner a procedure was developed which used both transient and steady-state system data to generate a steady-state locally-linear model analogous to (1-4). From this linear model, and from similar models for constrained functions $G(\underline{u})$, procedures were developed for selecting the adjustment vector $\Delta\underline{u}$ of (1-6). A cyclic algorithm was thus developed, with set point adjustment following linear model generation and linear model generation following set point adjustment.

Outline of the Thesis

The details of the development of the cyclic control algorithm are given in Chapters II through V. The systems which were considered are described in Chapter II, and the new system model, the essential feature of which is a detailed representation of system transport delays, is given in Chapter III. The extremum control algorithm is developed in Chapter IV for the unconstrained case and extended in Chapter V to the constrained case.

The remaining chapters are devoted to an analysis of the properties of the algorithm. Chapter VI considers the effects of measurement noise and Chapter VII considers the hardware required to implement the algorithm. Digital simulation is used in Chapter VIII to investigate the operational characteristics of the algorithm, with example systems including both tubular and stirred-tank reactors. Finally, conclusions and recommendations are given in Chapter IX.

CHAPTER II

DEFINITION OF THE CONTROL PROBLEM

The class of systems which were considered are described by first-order vector partial differential equations. Also considered was a degenerate class of systems described by ordinary differential equations. These systems were assumed to be controlled by set point controllers and to be constrained to satisfy magnitude constraints on the controller set points and on certain monitored system variables. The extremum control algorithm was required to adjust the set points, subject to these constraints, to maximize a given function of monitored system states.

Basic System Equations

The class of systems which were considered are described by the vector partial differential equation

$$\frac{\partial \underline{x}(\ell,t)}{\partial t} + v(\ell,t)\,\frac{\partial \underline{x}(\ell,t)}{\partial \ell} = \underline{f}[\underline{x}(\ell,t),u(\ell,t),\ell,t]. \qquad (2\text{-}1)$$

This equation describes the behavior of the states $\underline{x}(\ell,t)$ of a material flowing in the ℓ direction with a velocity $v(\ell,t)$. More specifically,

$$\underline{x}(\ell,t) = \begin{bmatrix} x_1(\ell,t) \\ \vdots \\ x_m(\ell,t) \end{bmatrix} = \text{vector of system states,} \qquad (2\text{-}2)$$

$$\underline{f}[\cdot] = \begin{bmatrix} f_1[\cdot] \\ \vdots \\ f_m[\cdot] \end{bmatrix} = \text{vector of continuous functions,} \qquad (2\text{-}3)$$

and

$v(\ell,t)$ = fluid velocity in direction of increasing ℓ (2-4)

$u(\ell,t)$ = scalar distributed control

ℓ = spatial variable, defined on $0<\ell\leq L$

t = time, defined on $t\geq 0$.

Associated with the system partial differential equations were initial conditions

$$\underline{x}(\ell,0) = \underline{x}_I(\ell) \qquad (2\text{-}5)$$

and boundary conditions

$$\underline{x}(0,t) = \underline{x}_B(t). \qquad (2\text{-}6)$$

The systems considered were n-stage systems for which the distributed control $u(\ell,t)$ was given by

$$u(\ell,t) = \begin{cases} u_j(t) & L_{j*}<\ell\leq L_j \\ & \qquad j=1,\ldots,n \\ 0 & L_j<\ell\leq L_j^* \end{cases} \tag{2-7}$$

where the jth stage was defined as $L_{j*}<\ell\leq L_j^*$. The inlet of the first stage was the system inlet,

$$L_{1*} = 0, \tag{2-8}$$

and the exit of the last stage was the system exit,

$$L_n^* = L. \tag{2-9}$$

The exit of the jth stage was the inlet of the (j+1)st,

$$L_j^* = L_{j+1*}. \tag{2-10}$$

This class of systems includes several configurations of multi-jacketed tubular reactors[4,18,33] and sequences of adiabatic catalytic reactors with heat exchangers between stages,[9,34] schematic diagrams of which are given in Appendix A. In these systems, the states $\underline{x}(\ell,t)$ might be the concentrations and temperature of the reacting species and the control $u(\ell,t)$ the temperature of a coolant in the surrounding cooling jackets.

This class of systems is inherently stable,[33] with all

disturbances propagating out of the system in a time less than or equal to

$$\int_0^L \frac{1}{v(\ell,t)}\, d\ell. \tag{2-11}$$

It will be required in the subsequent development that the systems satisfy certain sensitivity conditions. These conditions will be given at the appropriate point.

Degenerate System Equations

Also considered was a class of systems for which one or more of the n stages was described by a degenerate form of (2-1)-(2-7), namely, where the interval $L_{j*}<\ell\leq L_j$ contained only the point $\ell = L_j$ and a lumped-parameter sub-system described by

$$\frac{d\underline{x}(L_j,t)}{dt} = \underline{f}^j[\underline{x}(L_j,t),\underline{x}(L_{j*},t),u_j(t),t] \tag{2-12}$$

existed at that point.

Consideration of this degenerate form permitted results to be extended to stirred-tank reactors[16,19,34] and to the series combination of tubular and stirred-tank reactors,[11,35] schematic diagrams of which are given in Appendix A. In these systems, the states $\underline{x}(L_j,t)$ might be the concentrations and temperature of the mixture in the tank, the states $\underline{x}(L_{j*},t)$ the corresponding states of the stream entering the tank, and the control $u_j(t)$ the temperature of a coolant in a cooling coil.

It is necessary to require that this degenerate class of systems be stable, for it does not have the inherent stability of the basic system.[36,37] Both the required stability condition and a required sensitivity condition will be given at the appropriate point in the development.

Set Point Controller Equations

The controls $u_j(t)$ were assumed to be regulated by set point controllers to which were applied, at discrete times t_{ja}, a sequence of set points u_{ja} where the index a=1,2,... denoted the successive algorithmic cycles. Both the t_{ja} and the u_{ja} were specified by the extremum control algorithm.

Of interest in the research were ideal set point controllers, which were defined as ones which maintained $u_j(t)$ at u_{ja} from the command time t_{ja} to the next command time $t_{j,a+1}$,

$$u_j(t) = u_{ja} \qquad t_{ja} \leq t < t_{j,a+1} \tag{2-13}$$

$$j=1,\ldots,n \qquad a=1,2,\ldots,$$

and real set point controllers, which were defined as ones which required a finite time $\tau_{jc} \geq 0$ to achieve the set point u_{ja} and then only maintained $u_j(t)$ within a range $\delta_{ju} \geq 0$ of u_{ja},

$$|u_j(t)-u_{ja}| \leq \delta_{ju} \qquad t_{ja}+\tau_{jc} \leq t < t_{j,a+1} \tag{2-14}$$

$$j=1,\ldots,n \qquad a=1,2,\ldots .$$

The ideal set point controllers were a special case of the real set point controllers obtained by setting $\tau_{jc}=0$ and $\delta_{ju}=0$. A sketch of the response of a real set point controller is given in Figure 3.

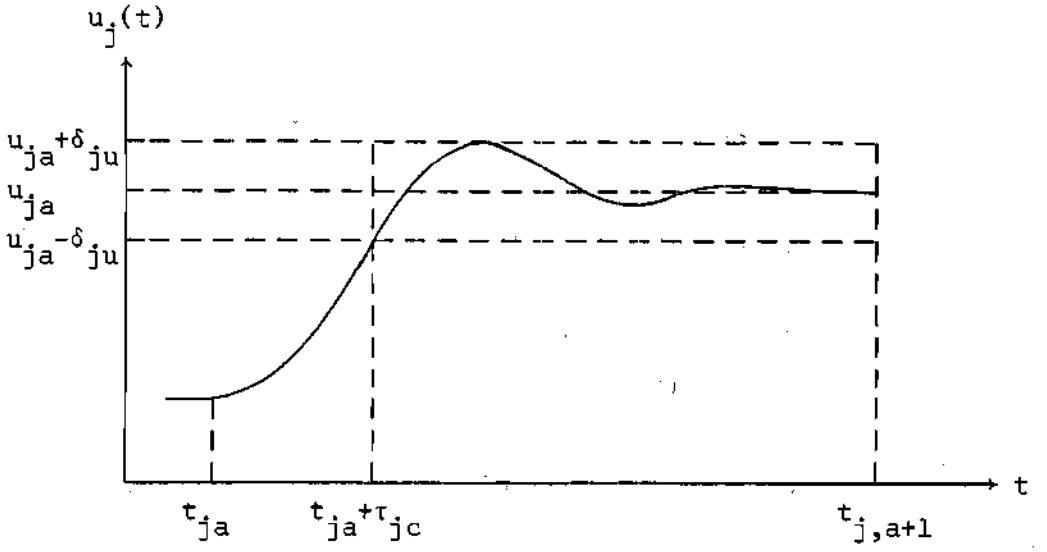


Figure 3. Response of a Real Set Point Controller

System Constraint Equations

Two types of constraints were imposed on the systems considered. The first was that the controls $u_j(t)$ should lie between prescribed lower and upper bounds u_{j*} and u_j^*:

$$u_{j*} \le u_j(t) \le u_j^* \qquad j=1,\ldots,n. \tag{2-15}$$

The second was that monitored vector functions $\underline{z}_{jk}$ of the states $\underline{x}(\ell_{jk},t)$ should lie between prescribed lower and upper bounds $\underline{z}_{jk*}$ and $\underline{z}_{jk}^*$. The $\underline{z}_{jk}[\underline{x}(\ell_{jk},t)]$ were obtained from measuring instruments located at points $\ell = \ell_{jk}$ in the jth stage,

$$L_{j*} < \ell_{jk} \leq L_j^* \qquad k=1,\ldots,p_j, \tag{2-16}$$

where the index "k" denoted the particular point in the jth stage and the integer "p_j" was the number of such points in the jth stage. The vector $\underline{z}_{jk}$ had q_{jk} components,

$$\underline{z}_{jk}[\underline{x}(\ell_{jk},t)] = \begin{bmatrix} z_{jk1}[\underline{x}(\ell_{jk},t)] \\ \vdots \\ z_{jkq_{jk}}[\underline{x}(\ell_{jk},t)] \end{bmatrix}, \tag{2-17}$$

on which were imposed the constraints

$$z_{jkr*} \leq z_{jkr}[\underline{x}(\ell_{jk},t)] \leq z_{jkr}^* \tag{2-18}$$

$$j=1,\ldots,n \qquad k=1,\ldots,p_j \qquad r=1,\ldots,q_{jk}.$$

In the systems illustrated in Appendix A, the constrained $\underline{z}_{jk}[\underline{x}(\ell_{jk},t)]$ might be monitored reactant temperatures and/or exit concentrations.[18,38]

Extremum Control Objective

The extremum control algorithm was to adjust the set points u_{ja} to maximize a function of the exit states $\underline{x}(L,t)$ and the set points themselves. Information on the exit states, obtained from a measuring instrument located at $\ell=L$, was available in the form of a measurement vector $\underline{h}[\underline{x}(L,t)]$, where

$$\underline{h}[\underline{x}(L,t)] = \begin{bmatrix} h_1[\underline{x}(L,t)] \\ \vdots \\ h_\mu[\underline{x}(L,t)] \end{bmatrix}. \qquad (2\text{-}19)$$

The measuring instrument was assumed to be of the "sample and analyze" type with an analysis time $\tau_m \geq 0$.[39,40] The vector $\underline{h}[\underline{x}(L,t)]$ was available only for discrete times $t=t^{\alpha}$ where the index $\alpha=1,2,\ldots$ denoted the successive measurement sample times. The function to be maximized was of the form

$$F(u_{1a},\ldots,u_{na},t^a) = G\left[\underline{h}[\underline{x}(L,t^a)],u_{1a},\ldots,u_{na}\right] \qquad (2\text{-}20)$$

where G was a known function with continuous first partial derivatives in its arguments $\underline{h}$ and $u_{1a},\ldots,u_{na}$. The set of t^a was a subset of the t^{α}. One t^{α} per algorithmic cycle was defined as t^a. Thus, the extremum control algorithm had to be organized so that

$$t^{\alpha+1} \geq t^{\alpha} + \tau_m \qquad \alpha=1,2,\ldots, \qquad (2\text{-}21)$$

and also so that any control response depending on the measurement $\underline{h}[\underline{x}(L,t^{\alpha})]$ would not be required until a time $t \geq t^{\alpha} + \tau_m$. In the systems illustrated in Appendix A, the function to be maximized is generally the exit concentration of the species being manufactured.[18,19]

Concluding Remarks

The solution of the problem outlined in this chapter required the coordination and control of the n system stages, the n set point controllers, and the many measuring instruments associated with the monitored functions $\underline{z}_{jk}[\underline{x}(\ell_{jk},t)]$ and $\underline{h}[\underline{x}(L,t)]$. These system elements, along with the process-control computer which was to implement the control algorithm, are shown in Figure 4. Scalar control and information paths are shown as single lines, and vector paths as double lines. The spatial relations within the respective system stages may be seen in the schematic diagrams of Appendix A.

The extremum control algorithm was to adjust the set points u_{ja} to maximize the function $F(u_{1a},\ldots,u_{na},t^a)$ subject to the constraints on the $u_j(t)$ and the $\underline{z}_{jk}[\underline{x}(\ell_{jk},t)]$. The first step in establishing this control algorithm was to develop a system model which adequately represented the relation between the set points u_{ja} and the monitored functions $\underline{z}_{jk}[\underline{x}(\ell_{jk},t)]$ and $\underline{h}[\underline{x}(L,t)]$. This development is given in the following chapter.

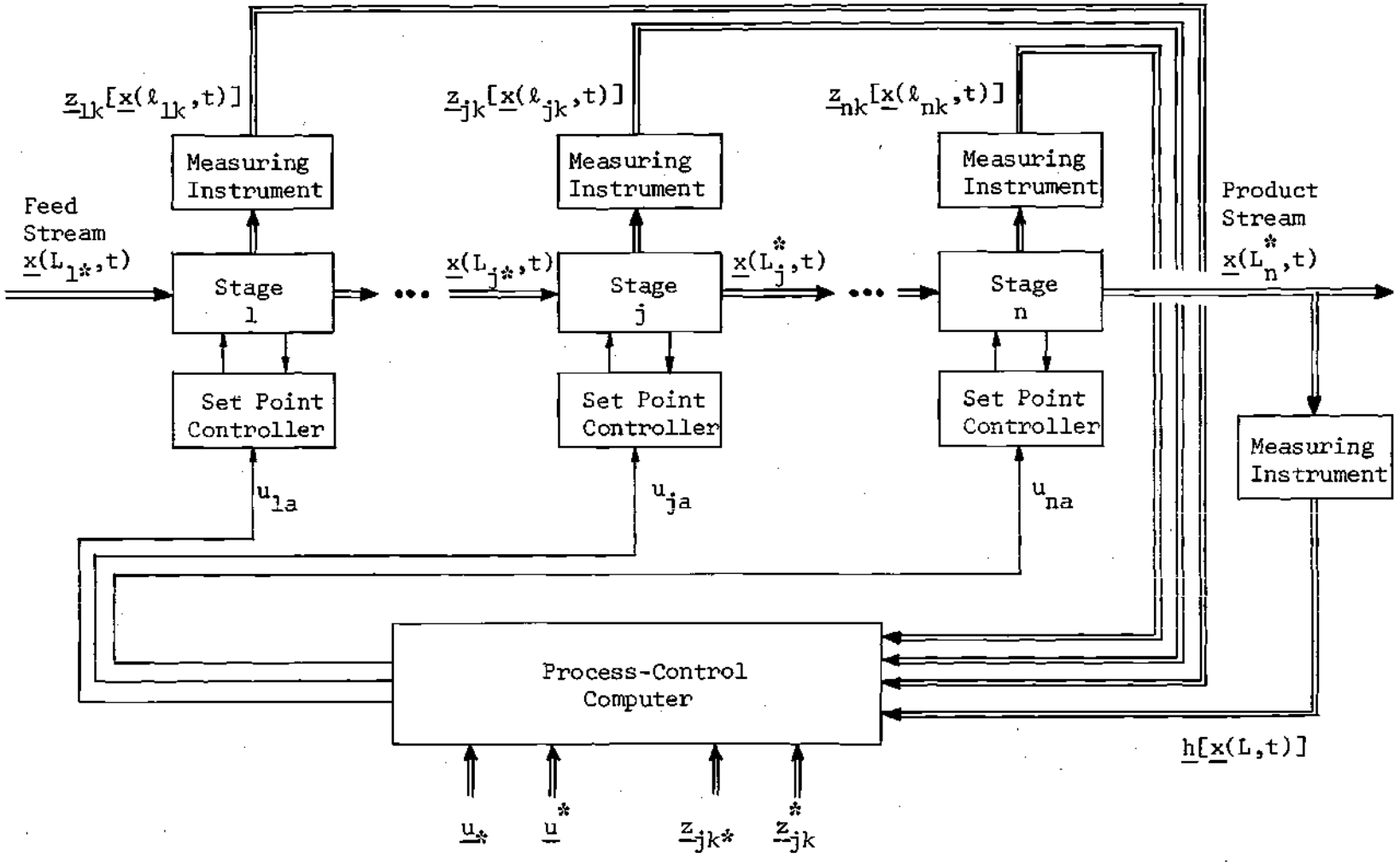


Figure 4. Multistage System with Computer Interconnections

CHAPTER III

DEVELOPMENT OF THE SYSTEM MODEL

The system model was developed by transforming the partial differential equations (2-1) into a set of ordinary differential equations. These equations were then solved on $L_{j*} < \ell \leq L_j^*$ and the solution used to establish a model on the jth system stage. This model was then coupled to the corresponding model of the (j-1)st stage, and the complete system model established by continuing the coupling until the 1st stage was included.

Method of Characteristics

Central to the development of the model is the method of characteristics,[41] a procedure for transforming a partial differential equation into a pair of ordinary differential equations. Consider, in this regard, the lines $t(\ell;t_\alpha)$ in the ℓ-t plane which are solutions to the ordinary differential equation

$$\frac{dt(\ell)}{d\ell} = \frac{1}{v(\ell,t)} \qquad 0<\ell\leq L \tag{3-1}$$

with initial condition

$$t(0) = t_\alpha \geq 0. \tag{3-2}$$

Each initial condition t_α defines a solution $t(\ell;t_\alpha)$ which is simply the time at which a fluid element entering the system at $t=t_\alpha$ reaches the spatial coordinate ℓ. Two such solutions are shown in Figure 5.

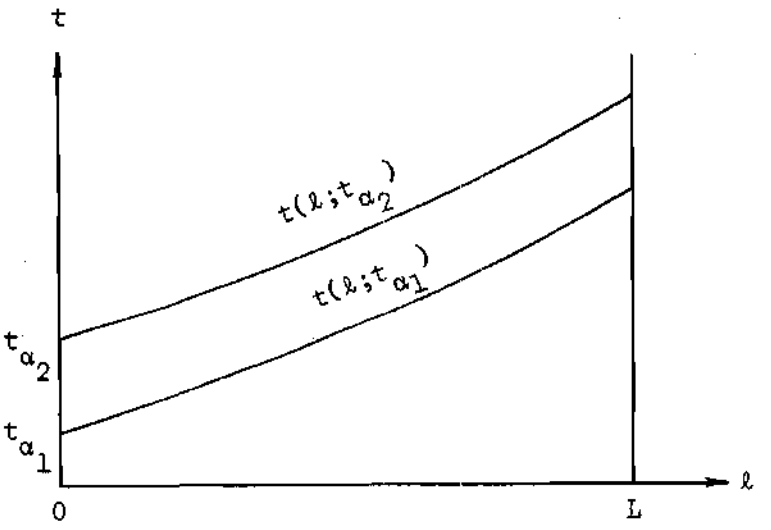


Figure 5. Characteristic Lines in the ℓ-t Plane

Now consider the transformation of (2-1) into an ordinary differential equation. Along the characteristic lines $t(\ell;t_\alpha)$, the total derivative of $\underline{x}(\ell,t)$ with respect to ℓ is

$$\frac{d\underline{x}(\ell,t)}{d\ell} = \frac{\partial \underline{x}(\ell,t)}{\partial t}\frac{dt(\ell)}{d\ell} + \frac{\partial \underline{x}(\ell,t)}{\partial \ell} \qquad (3\text{-}3)$$

$$= \frac{\partial \underline{x}(\ell,t)}{\partial t}\frac{1}{v(\ell,t)} + \frac{\partial \underline{x}(\ell,t)}{\partial \ell} .$$

Combining (2-1) and (3-3),

$$\frac{d\underline{x}(\ell,t)}{d\ell} = \frac{1}{v(\ell,t)}\,\underline{f}[\underline{x}(\ell,t),u(\ell,t),\ell,t] \qquad 0<\ell\leq L. \tag{3-4}$$

This equation is an ordinary differential equation in the independent variable ℓ with t as a parameter. The t corresponding to any particular ℓ is given by $t(\ell;t_\alpha)$.

To solve (3-4), it is necessary to establish a boundary condition on $\underline{x}(\ell,t)$. To this end, define the function $\tau(\ell,t)$ to be the time required for a fluid element reaching the point ℓ at a time t to traverse the distance $|\ell-0|$:

$$\tau(\ell,t) = t(\ell;t_\alpha) - t_\alpha. \tag{3-5}$$

Recalling that t in (3-4) is a parameter whose value depends on ℓ, it follows from (2-6) that the boundary condition associated with (3-4) is

$$\underline{x}\big(0,t-\tau(\ell,t)\big) = \underline{x}_B\big(t-\tau(\ell,t)\big). \tag{3-6}$$

The solution to (2-1) along the characteristic lines $t(\ell,t_\alpha)$ may be obtained by solving simultaneously the ordinary differential equations (3-1) and (3-4), utilizing their associated initial and boundary conditions (3-2) and (3-6). Intuitively speaking, solution by the method of characteristics amounts to following successive fluid elements from ℓ=0 to ℓ=L. The behavior of each element is governed by an *ordinary* differential equation because the element state depends only on its interaction with the stationary system.

Model of Basic jth Stage

Consider now only the jth stage of the basic system, and in particular the interval $L_{j*} < \ell \leq L_j$. Inserting (2-7) into (3-4),

$$\frac{d\underline{x}(\ell,t)}{d\ell} = \frac{1}{v(\ell,t)}\,\underline{f}[\underline{x}(\ell,t),u_j(t),\ell,t] \qquad L_{j*}<\ell\leq L_j. \tag{3-7}$$

Boundary Condition

To establish the boundary condition for (3-7), define the function $\sigma_j(\ell,t)$ to be the time required for a fluid element reaching the point ℓ at a time t to traverse the distance $|\ell-L_{j*}|$:

$$\sigma_j(\ell,t) = t(\ell;t_\alpha) - t(L_{j*};t_\alpha). \tag{3-8}$$

Then, analogous to (3-6), the boundary condition for (3-7) is

$$\underline{x}\left(L_{j*},t-\sigma_j(\ell,t)\right). \tag{3-9}$$

To permit later coupling of the jth stage model with the corresponding (j-1)st stage model, impose on (3-9) the dummy restriction

$$s_{j*} \leq t \leq s_j^*. \tag{3-10}$$

This restriction may be thought of as specifying the time interval in which the stage inlet states $\underline{x}(L_{j*},t)$ are at steady state.

Controller Assumption

Now assume that the δ_{ju} of (2-14) are small enough that fluctuations in $u_j(t)$ about u_{ja} may be neglected for $t_{ja} + \tau_{jc} \leq t < t_{j,a+1}$. That is, assume that

$$u_j(t) = u_{ja} \qquad t_{ja}+\tau_{jc} \leq t < t_{j,a+1} \tag{3-11}$$

$$j=1,\ldots,n \qquad a=1,2,\ldots .$$

Proper definition of τ_{jc}, of course, is the key to the validity of (3-11). Now inserting (3-11) into (3-7),

$$\frac{d\underline{x}(\ell,t)}{d\ell} = \frac{1}{v(\ell,t)}\,\underline{f}[\underline{x}(\ell,t),u_{ja},\ell,t] \qquad L_{j*} < \ell \leq L_j . \tag{3-12}$$

Basic Solution

The solution to (3-12) is a function of ℓ and is dependent on the boundary condition $\underline{x}\left(L_{j*},t-\sigma_j(\ell,t)\right)$ and the parameters u_{ja} and t: [42]

$$\underline{x}(\ell,t) = \underline{S}_j[\ell;\underline{x}\left(L_{j*},t-\sigma_j(\ell,t)\right),u_{ja},t] \qquad L_{j*} < \ell \leq L_j . \tag{3-13}$$

Validity Conditions

Two conditions are imposed on the validity of (3-13). The first, shown schematically in Figure 6, arises from the stage inlet restriction (3-10). The first fluid element satisfying (3-10) reaches the point ℓ at

$$t = s_{j*} + \sigma_j(\ell,t) \tag{3-14}$$

and the last fluid element satisfying (3-10) reaches ℓ at

$$t = s_j^* + \sigma_j(\ell,t). \tag{3-15}$$

Combining (3-14) and (3-15), the validity condition

$$s_{j*} + \sigma_j(\ell,t) \le t \le s_j^* + \sigma_j(\ell,t) \tag{3-16}$$

is obtained. The second validity condition, shown schematically in Figure 7, arises from the replacement of $u_j(t)$ by u_{ja} in (3-12). This replacement is valid for fluid elements which encounter u_{ja} throughout their traversal of the distance $|\ell - L_{j*}|$. The first fluid element satisfying this condition reaches ℓ at

$$t = t_{ja} + \tau_{jc} + \sigma_j(\ell,t) \tag{3-17}$$

and the last fluid element satisfying this condition reaches ℓ at

$$t = t_{j,a+1}. \tag{3-18}$$

Combining (3-17) and (3-18), the second validity condition is obtained:

$$t_{ja} + \tau_{jc} + \sigma_j(\ell,t) \le t \le t_{j,a+1}. \tag{3-19}$$

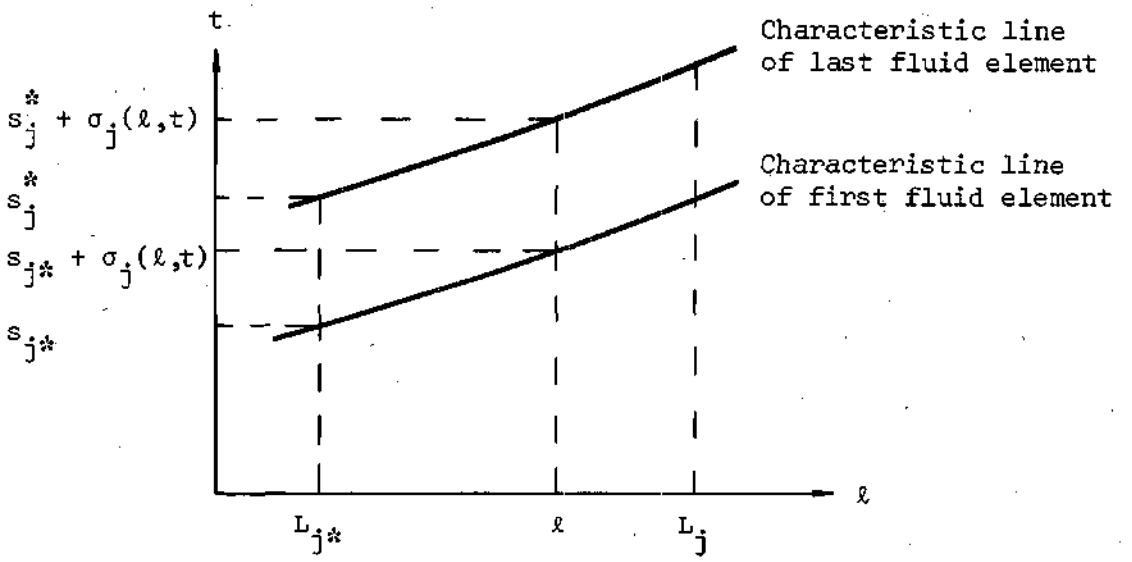


Figure 6. First Validity Condition for Basic Solution

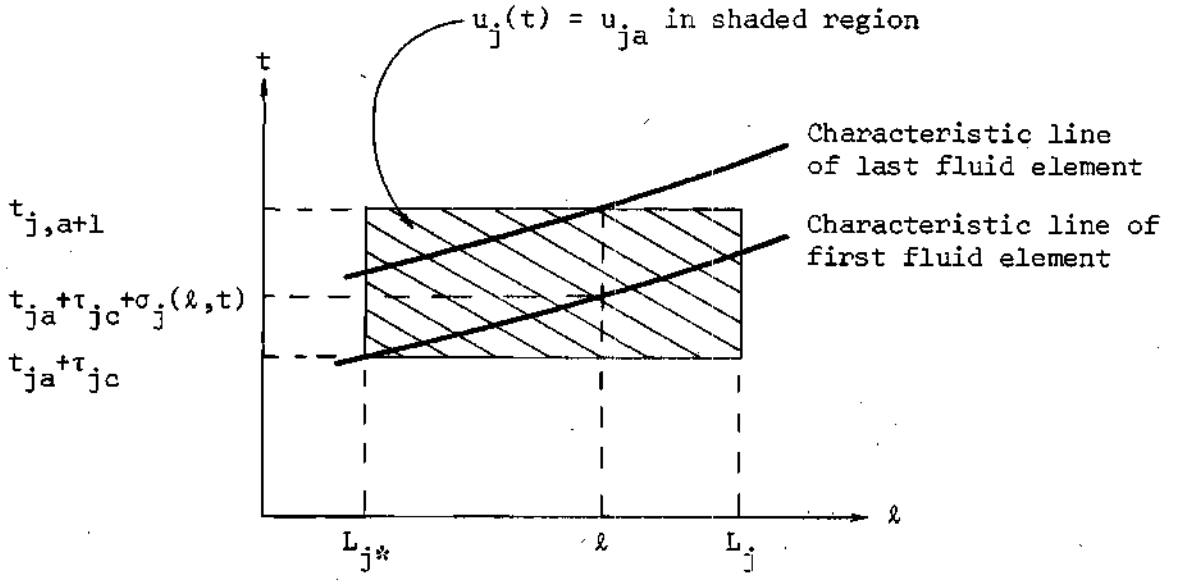


Figure 7. Second Validity Condition for Basic Solution

Combining (3-16) and (3-19), the single condition

$$\text{Max}\{t_{ja}+\tau_{jc}+\sigma_j(\ell,t), s_{j*}+\sigma_j(\ell,t)\} \tag{3-20}$$

$$\leq t \leq \text{Min}\{t_{j,a+1}, s_j^*+\sigma_j(\ell,t)\}$$

is obtained, where the functions Max{·} and Min{·} are, respectively, the maximum and minimum values of their several arguments.

Stage Secondary

Consider now that portion of the jth stage defined on $L_j<\ell\leq L_j^*$. Inserting (2-7) into (3-4),

$$\frac{d\underline{x}(\ell,t)}{d\ell} = \frac{1}{v(\ell,t)}\,\underline{f}[\underline{x}(\ell,t),0,\ell,t] \qquad L_j<\ell\leq L_j^*. \tag{3-21}$$

Defining $\rho_j(\ell,t)$ as the time required for a fluid element reaching the point ℓ at a time t to traverse the distance $|\ell-L_j|$,

$$\rho_j(\ell,t) = t(\ell;t_\alpha) - t(L_j;t_\alpha), \tag{3-22}$$

the boundary condition

$$\underline{x}\big(L_j, t-\rho_j(\ell,t)\big) \tag{3-23}$$

is obtained.

Secondary Solution

The solution to (3-21) is a function of ℓ and is dependent on the boundary condition $\underline{x}\left(L_j, t-\rho_j(\ell,t)\right)$ and the parameter t: [42]

$$\underline{x}(\ell,t) = \underline{T}_j[\ell;\underline{x}\left(L_j, t-\rho_j(\ell,t)\right),t] \quad L_j<\ell\leq L_j^*. \tag{3-24}$$

It follows from (3-13) that the boundary condition in (3-24) is given by

$$\underline{x}\left(L_j, t-\rho_j(\ell,t)\right) = \underline{S}_j\Big[L_j;\underline{x}\Big(L_{j*}, t-\rho_j(\ell,t) - \sigma_j\left(L_j, t-\rho_j(\ell,t)\right)\Big), u_{ja}, t-\rho_j(\ell,t)\Big]. \tag{3-25}$$

But since

$$\sigma_j\left(L_j, t-\rho_j(\ell,t)\right) + \rho_j(\ell,t) = \sigma_j(\ell,t), \tag{3-26}$$

(3-25) may be simplified to

$$\underline{x}\left(L_j, t-\rho_j(\ell,t)\right) = \underline{S}_j\left[L_j;\underline{x}\left(L_{j*}, t-\sigma_j(\ell,t)\right), u_{ja}, t-\rho_j(\ell,t)\right]. \tag{3-27}$$

Equation (3-27) may then be substituted into (3-24) to yield

$$\underline{x}(\ell,t) = \underline{T}_j\left[\ell;\underline{S}_j\left[L_j;\underline{x}\left(L_{j*}, t-\sigma_j(\ell,t)\right), u_{ja}, t-\rho_j(\ell,t)\right],t\right] \quad L_j<\ell\leq L_j^*. \tag{3-28}$$

This expression is simply a function of ℓ dependent on the boundary condition $\underline{x}\left(L_{j*}, t-\sigma_j(\ell,t)\right)$ and the parameters u_{ja} and t, and may be simplified by defining its right-hand side to be a new function $\underline{W}_j$:

$$\underline{x}(\ell,t) = \underline{W}_j\left[\ell; \underline{x}\left(L_{j*}, t-\sigma_j(\ell,t)\right), u_{ja}, t\right] \quad L_j < \ell \leq L_j^*. \tag{3-29}$$

Secondary Validity Conditions

The validity conditions imposed on (3-29) may be obtained in the same manner as those imposed on (3-13). Equations (3-14)-(3-16), in fact, are still valid and are shown schematically in Figure 8. Equation (3-17) is also valid, but (3-18) must be changed since the last fluid encountering u_{ja} throughout its traversal of the distance $|\ell - L_{j*}|$ reaches the point ℓ at

$$t = t_{j,a+1} + \rho_j(\ell,t). \tag{3-30}$$

Equations (3-17) and (3-30) are shown schematically in Figure 9. These two equations may be combined to obtain the condition analogous to (3-19):

$$t_{ja} + \tau_{jc} + \sigma_j(\ell,t) \leq t \leq t_{j,a+1} + \rho_j(\ell,t). \tag{3-31}$$

Then combining (3-16) and (3-31),

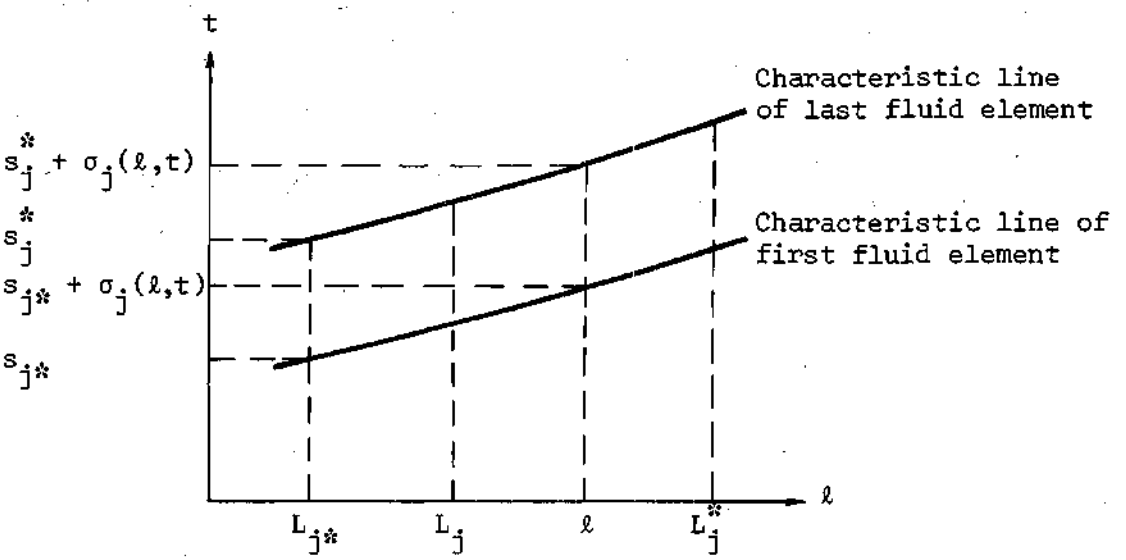


Figure 8. First Validity Condition for Secondary Solution

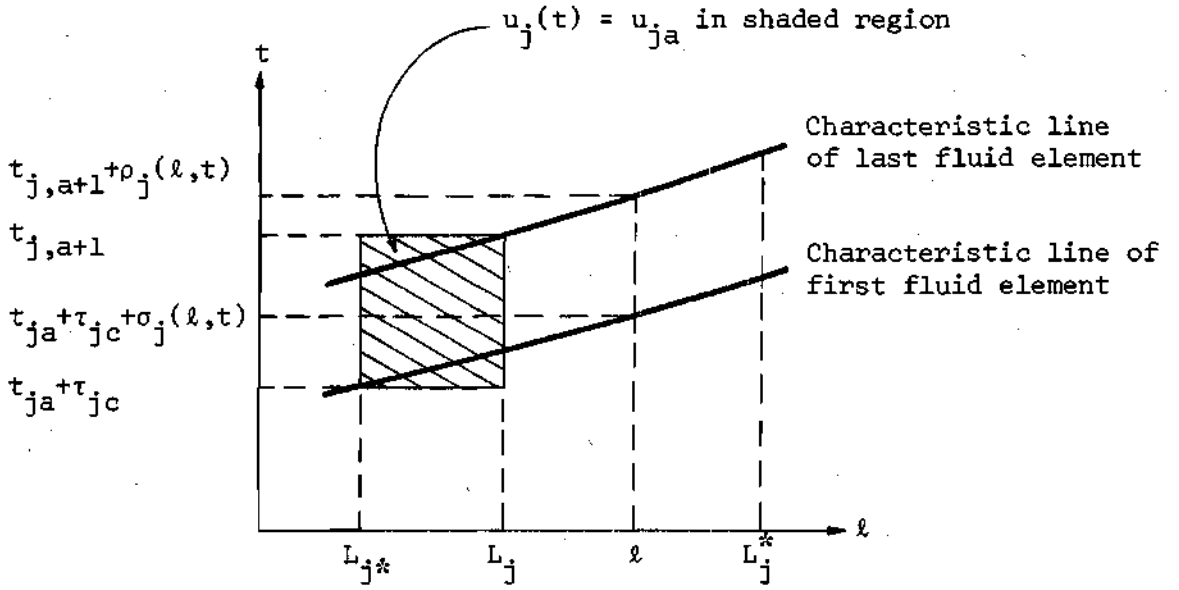


Figure 9. Second Validity Condition for Secondary Solution

$$\text{Max}\{t_{ja}+\tau_{jc}+\sigma_j(\ell,t),\ s_{j*}+\sigma_j(\ell,t)\} \le t \qquad (3\text{-}32)$$

$$\le \text{Min}\{t_{j,a+1}+\rho_j(\ell,t),\ s_j^*+\sigma_j(\ell,t)\}.$$

Model of Degenerate jth Stage

Consider now the lumped-parameter sub-system existing at $\ell=L_j$ and defined by (2-12):

$$\frac{d\underline{x}(L_j,t)}{dt} = \underline{f}^j[\underline{x}(L_j,t),\ \underline{x}(L_{j*},t),\ u_j(t),\ t]. \qquad (3\text{-}33)$$

Analogous to (3-10), let the stage inlet states $\underline{x}(L_{j*},t)$ be at steady state in the interval

$$s_{j*} \le t \le s_j^*. \qquad (3\text{-}34)$$

And, let the assumption (3-11) apply also to the degenerate stage, from which it follows that (3-33) may be written

$$\frac{d\underline{x}(L_j,t)}{dt} = \underline{f}^j[\underline{x}(L_j,t),\ \underline{x}(L_{j*},t),\ u_{ja},\ t]. \qquad (3\text{-}35)$$

The validity condition imposed on (3-35) may be obtained by combining (3-11) and (3-34):

$$\text{Max}\{t_{ja}+\tau_{jc},\ s_{j*}\} \le t < \text{Min}\{t_{j,a+1},\ s_j^*\}. \qquad (3\text{-}36)$$

The initial condition associated with (3-35) is

$$\underline{x}(L_j,\ \text{Max}\{t_{ja}+\tau_{jc}, s_{j*}\}). \tag{3-37}$$

Degenerate Solution

The solution to (3-35) is a function of t and is dependent on the initial condition $\underline{x}(L_j,\ \text{Max}\{t_{ja}+\tau_{jc}, s_{j*}\})$ and the parameters $\underline{x}(L_{j*},t)$ and u_{ja}:[42]

$$\underline{x}(L_j,t) = \underline{S}_j'[t;\underline{x}(L_j,\text{Max}\{t_{ja}+\tau_{jc}, s_{j*}\}),\underline{x}(L_{j*},t),u_{ja}]. \tag{3-38}$$

It is necessary at this point to impose a stability condition on the degenerate stage. This condition, rather loosely stated, is that after a time $\tau_{js}\geq 0$ the system should have settled to a constant, or slowly varying, state which is dependent on $\underline{x}(L_{j*},t)$, u_{ja}, and t, but not on the stage initial condition $\underline{x}(L_j,\text{Max}\{t_{ja}+\tau_{jc}, s_{j*}\})$:

$$\underline{x}(L_j,t) = \underline{S}_j''[\underline{x}(L_{j*},t),u_{ja},t]. \tag{3-39}$$

As was the case with (3-11), proper definition of τ_{js} is the key to the validity of (3-39).

Validity Conditions

The validity condition imposed on (3-39) may be obtained from (3-36) by adding τ_{js} to the lower bound on t:

$$\text{Max}\{t_{ja}+\tau_{jc}, s_{j*}\} + \tau_{js} \le t \le \text{Min}\{t_{j,a+1}, s_j^*\}. \qquad (3\text{-}40)$$

The upper bound on t was extended to equality since only the solution to (3-35), as opposed to (3-35) itself, was under consideration here.

Secondary Solution

That portion of the degenerate stage defined on $L_j < \ell \le L_j^*$ is identical to the corresponding portion of the basic stage. It follows, then, that (3-24) is still valid for the degenerate stage:

$$\underline{x}(\ell,t) = \underline{T}_j\left[\ell; \underline{x}\left(L_j, t-\rho_j(\ell,t)\right), t\right] \quad L_j < \ell \le L_j^*. \qquad (3\text{-}41)$$

The $\underline{x}\left(L_j, t-\rho_j(\ell,t)\right)$ to be inserted into (3-41), however, is not obtained from the basic stage analysis but rather from the degenerate stage analysis. Specifically, from (3-39),

$$\underline{x}\left(L_j, t-\rho_j(\ell,t)\right) = \underline{S}_j''\left[\underline{x}\left(L_{j*}, t-\rho_j(\ell,t)\right), u_{ja}, t-\rho_j(\ell,t)\right]. \qquad (3\text{-}42)$$

Inserting (3-42) into (3-41),

$$\underline{x}(\ell,t) = \underline{T}_j\left[\ell; \underline{S}_j''\left[\underline{x}\left(L_{j*}, t-\rho_j(\ell,t)\right), u_{ja}, t-\rho_j(\ell,t)\right], t\right] \qquad (3\text{-}43)$$

$$L_j < \ell \le L_j^*.$$

This equation may be simplified by defining its right-hand side to be a new function $\underline{W}_j'$:

$$\underline{x}(\ell,t) = \underline{W}_j'\left[\ell;\underline{x}\left(L_{j*},t-\rho_j(\ell,t)\right),u_{ja},t\right] \quad L_j<\ell\le L_j^*. \tag{3-44}$$

Secondary Validity Conditions

The validity conditions imposed on (3-44) may be obtained from (3-40) by replacing t by $t - \rho_j(\ell,t)$:

$$\text{Max}\{t_{ja}+\tau_{jc},s_{j*}\} + \tau_{js} \le t - \rho_j(\ell,t) \le \text{Min}\{t_{j,a+1},s_j^*\}. \tag{3-45}$$

This inequality may then be rearranged to give

$$\text{Max}\{t_{ja}+\tau_{jc},s_{j*}\} + \tau_{js} + \rho_j(\ell,t) \le t \tag{3-46}$$

$$\le \text{Min}\{t_{j,a+1},s_j^*\} + \rho_j(\ell,t).$$

This and other validity conditions, it should be recalled, simply specify the times during which the states $\underline{x}(\ell,t)$ are at steady state.

Model of General jth Stage

The preceding analyses have established solution equations and corresponding validity conditions for four distinct cases:

Basic jth stage	$L_{j*} < \ell \le L_j$
Basic jth stage	$L_j < \ell \le L_j^*$
Degenerate jth stage	$\ell = L_j$
Degenerate jth stage	$L_j < \ell \le L_j^*$.

The solution equations are given, respectively, by (3-13), (3-29),

(3-39), and (3-44), and the corresponding validity conditions by (3-20), (3-32), (3-40), and (3-46).

Appropriate definitions of the functions $\underline{X}_j[\ell;\underline{x},u,t]$ and $\bar{\sigma}_j(\ell,t)$, given in Appendix B, permit the four solution equations to be represented by the single equation

$$\underline{x}(\ell,t) = \underline{X}_j\left[\ell;\underline{x}\left(L_{j*},t-\bar{\sigma}_j(\ell,t)\right),u_{ja},t\right] \quad L_{j*}<\ell\leq L_j^*. \tag{3-47}$$

Similar definitions permit the four validity conditions to be represented by the single inequality

$$\text{Max}\{t_{ja}+\rho_j^*(\ell),s_{j*}+\sigma_j^*(\ell)\} \leq t \tag{3-48}$$

$$\leq \text{Min}\{t_{j,a+1}+\rho_{j*}(\ell),s_j^*+\sigma_{j*}(\ell)\}.$$

The derivation of (3-48) depends on the assumption that the fluid velocity $v(\ell,t)$ lies between known time-independent bounds $v_*(\ell)$ and $v^*(\ell)$:

$$v_*(\ell) \leq v(\ell,t) \leq v^*(\ell) \qquad 0<\ell\leq L. \tag{3-49}$$

Model of Complete System

The model of the complete system may be obtained by coupling the general jth stage model to the corresponding model of the (j-1)st stage, and continuing such coupling until the 1st stage is included.

The key to this coupling is the equality of the jth stage inlet states and the (j-1)st stage exit states. It follows from this equality that the $\underline{x}\left(L_{j*},t-\bar{\sigma}_j(\ell,t)\right)$ of (3-47) may be replaced by $\underline{x}\left(L^*_{j-1},t-\bar{\sigma}_j(\ell,t)\right)$. These states may then be modeled by the (j-1)st stage analog of (3-47) and the resulting model inserted back into (3-47). A similar procedure permits the s_{j*} and s^*_j of (3-48) to be established from the (j-1)st stage analog of (3-48) and the resulting expressions to be inserted back into (3-48).

The coupled solution equation, established by this procedure in Appendix C, is

$$\underline{x}(\ell,t) = \underline{Y}_j(\ell;u_{1a},\ldots,u_{ja},t) \qquad L_{j*}<\ell\leq L^*_j. \tag{3-50}$$

This equation models the states $\underline{x}(\ell,t)$ as functions of ℓ dependent on the parameters $u_{1a},\ldots,u_{ja}$ and t.

The validity condition associated with (3-50), also established in Appendix C, is

$$\begin{aligned}\text{Max}\{&t_{ja} + \rho^*_j(\ell),\ t_{j-1,a} + \rho^*_{j-1}(L^*_{j-1}) + \sigma^*_j(\ell), \\ &t_{j-2,a} + \rho^*_{j-2}(L^*_{j-2}) + \sigma^*_{j-1}(L^*_{j-1}) + \sigma^*_j(\ell),\ \ldots, \\ &t_{1a} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{j-1}(L^*_{j-1}) + \sigma^*_j(\ell)\} \\ \leq\ t \leq\ \text{Min}\{&t_{j,a+1} + \rho_{j*}(\ell),\ t_{j-1,a+1} + \rho_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell),\end{aligned} \tag{3-51}$$

$$t_{j-2,a+1} + \rho_{j-2*}(L^*_{j-2}) + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell), \ldots,$$

$$t_{1,a+1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell)\}.$$

This condition specifies the times at which the states $\underline{x}(\ell,t)$, as modeled by (3-50), are at steady state. The derivation of (3-51) depends on the assumption that the system boundary conditions $\underline{x}_B(t)$ are at steady state for all $t \geq 0$.

In comparing the complexity of (3-50) and (3-51), it should be evident that the essential feature of the system model is the detailed representation of system response times and transport delays. In subsequent chapters, this representation will be used to develop equations coordinating the control of the several system stages.

CHAPTER IV

DEVELOPMENT OF THE UNCONSTRAINED ALGORITHM

The system model given in the previous chapter was used to develop the n-stage unconstrained extremum control algorithm. Specifically, (3-50) was used to develop a locally-linear system representation analogous to (1-4), and (3-51) was used to coordinate the updating of a partial derivative approximation analogous to (1-5). A gradient technique was then used to select new system set points.

The procedure developed for updating the partial derivative approximation is the heart of the multiplexed gradient technique. This procedure interlaces the effects of several set point combinations by using system transport delays to separate fluid elements representative of different system states. These fluid elements, some representing transient states, are then sampled at appropriate times and the data obtained used to generate a steady-state system model.

Structure of the Algorithm

Consider first the function to be maximized, as given by (2-20):

$$F(u_{1a},\ldots,u_{na},t^a) = G\left[\underline{h}[\underline{x}(L,t^a)],u_{1a},\ldots,u_{na}\right]. \qquad (4\text{-}1)$$

The model of $\underline{x}(L,t^a)$, for insertion into (4-1), may be obtained from (3-50) by setting j to n, ℓ to L, and t to t^a:

$$\underline{x}(L,t^a) = \underline{Y}_n(L;u_{1a},\ldots,u_{na},t^a). \tag{4-2}$$

To simplify notation, define the function $\underline{H}$ by

$$\underline{H}(u_{1a},\ldots,u_{na},t^a) = \underline{h}[\underline{x}(L,t^a)]. \tag{4-3}$$

Then, combining (4-1)-(4-3),

$$F(u_{1a},\ldots,u_{na},t^a) = G[\underline{H}(u_{1a},\ldots,u_{na},t^a),u_{1a},\ldots,u_{na}]. \tag{4-4}$$

Locally-Linear Representation

Now let $\underline{H}$ be represented by a Taylor series expansion, truncated at the linear terms, in a neighborhood of the point $(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1})$:

$$\underline{H}(u_{1a},\ldots,u_{na},t^a) = \underline{H}(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}) \tag{4-5}$$

$$+ \sum_{j=1}^{n} \left.\frac{\partial \underline{H}}{\partial u_{ja}}\right|_{a-1} (u_{ja}-u_{j,a-1}) + \left.\frac{\partial \underline{H}}{\partial t^a}\right|_{a-1} (t^a-t^{a-1}).$$

This representation depends on the assumption that $\underline{H}$ has continuous first partial derivatives in its arguments $u_{1a},\ldots,u_{na}$ and t^a.

Now assume that the effect of temporal drift is small in comparison to the effect of set point changes:

$$\left|\left.\frac{\partial \underline{H}}{\partial t^a}\right|_{a-1}(t^a-t^{a-1})\right| << \left|\left.\frac{\partial \underline{H}}{\partial u_{ja}}\right|_{a-1}(u_{ja}-u_{j,a-1})\right| \qquad (4\text{-}6)$$

$$j=1,\ldots,n.$$

It then follows that (4-5) may be replaced by the simpler representation

$$\underline{H}(u_{1a},\ldots,u_{na},t^a) = \underline{H}(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}) \qquad (4\text{-}7)$$

$$+ \sum_{j=1}^{n} \left.\frac{\partial \underline{H}}{\partial u_{ja}}\right|_{a-1}(u_{ja}-u_{j,a-1}).$$

Now also let F be represented by a Taylor series expansion truncated at the linear terms:

$$F(u_{1a},\ldots,u_{na},t^a) = F(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}) \qquad (4\text{-}8)$$

$$+ \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}(u_{ja}-u_{j,a-1}).$$

The $\partial F/\partial t^a$ term is absent from this expansion because, from (4-4) and (4-7),

$$\left.\frac{\partial F}{\partial t^a}\right|_{a-1} = 0. \qquad (4\text{-}9)$$

Multiplexed Gradient Technique

The basic procedure used in the extremum control algorithm is to continually establish new set points u_{ja} from existing set points $u_{j,a-1}$ according to

$$u_{ja} = u_{j,a-1} + e_{ja}\Delta u_j \tag{4-10}$$

where the $e_{ja} = \pm 1$ are chosen to increase the value of $F(u_{1a},\ldots, u_{na},t^a)$:

$$e_{ja} = \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]. \tag{4-11}$$

The $\Delta u_j > 0$ are parameters set by the user to produce small (on the order of 1 per cent) changes in $F(u_{1a},\ldots,u_{na},t^a)$. The function sign[·] is the binary sign function

$$\text{sign}[x] = \begin{cases} +1 & x \geq 0 \\ -1 & x < 0. \end{cases} \tag{4-12}$$

Thus, the set points are changed by an amount Δu_j at each algorithmic cycle.

From (4-4), the partial derivatives of (4-11) are

$$\frac{\partial F}{\partial u_{ja}} = \frac{\partial G}{\partial H_1}\frac{\partial H_1}{\partial u_{ja}} + \ldots + \frac{\partial G}{\partial H_\mu}\frac{\partial H_\mu}{\partial u_{ja}} + \frac{\partial G}{\partial u_{ja}} . \qquad (4\text{-}13)$$

The partials of G with respect to $H_1,\ldots,H_\mu$ and u_{ja} may be obtained analytically from the known function $G[\underline{H},u_{1a},\ldots,u_{na}]$. The partials of $H_1,\ldots,H_\mu$ with respect to u_{ja}, however, must be approximated from measured system data. The approximation used is

$$\left.\frac{\partial \underline{H}}{\partial u_{ja}}\right|_{a-1} = \frac{1}{(u_{j,a-1}-u_{j,a-2})} \times \qquad (4\text{-}14)$$

$$[\underline{H}(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-1},u_{j+1,a-1},\ldots,u_{n,a-1},t^{j-1,a-1})$$

$$- \underline{H}(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-2},u_{j+1,a-1},\ldots,u_{n,a-1},t^{j,a-1})].$$

This approximation can easily be shown to be consistent with the linear representation (4-7).

The equations given above provide the basis for a cyclic extremum control algorithm: given changes in the set points at times $t_{j,a-1}$, measurements are made at times $t^{j,a-1}$ and (4-14) is used to approximate the partial derivatives of (4-13). New set points u_{ja} are then selected according to (4-10)-(4-12). The cycle is repeated continually with the set points first achieving and then tracking the drifting system optimum.

Philosophy of the Technique

It seems appropriate at this point to present the basic philosophy of the multiplexed gradient technique and, in particular, to provide an

interpretation of (4-14). In this regard, consider the simplest of all examples, a two-stage tubular reactor of length L with two jackets each of length L/2. Let the fluid velocity $v(\ell,t)$ be constant at the value v, and let the function to be maximized be the measured response H_1:

$$F(u_{1a},u_{2a},t^a) = H_1(u_{1a},u_{2a},t^a). \tag{4-15}$$

Let the system be at steady state with the $u_j(t)$ maintained at set points $u_{j,a-2}$ by ideal set point controllers. Then at $t_{j,a-1}$ let the set points be changed to $u_{j,a-1}$. The progress of three fluid elements passing through the system under these conditions is shown in Figure 10. It may be seen that the data required for the partial derivative approximations of (4-14), namely

$$\left.\frac{\partial H_1}{\partial u_{1a}}\right|_{a-1} = \frac{H_1(u_{1,a-1},u_{2,a-1},t^{0,a-1}) - H_1(u_{1,a-2},u_{2,a-1},t^{1,a-1})}{u_{1,a-1} - u_{1,a-2}} \tag{4-16}$$

$$\left.\frac{\partial H_1}{\partial u_{2a}}\right|_{a-1} = \frac{H_1(u_{1,a-2},u_{2,a-1},t^{1,a-1}) - H_1(u_{1,a-2},u_{2,a-2},t^{2,a-1})}{u_{2,a-1} - u_{2,a-2}},$$

may be obtained in the time required for one fluid element to traverse the length of the reactor. This means that the set points may be adjusted repeatedly at intervals of L/v time units, each time with the benefit of an updated locally-linear system model.

Another interpretation of the data required by (4-16) is given in Figure 11, where the measurements of Figure 10 are plotted on a

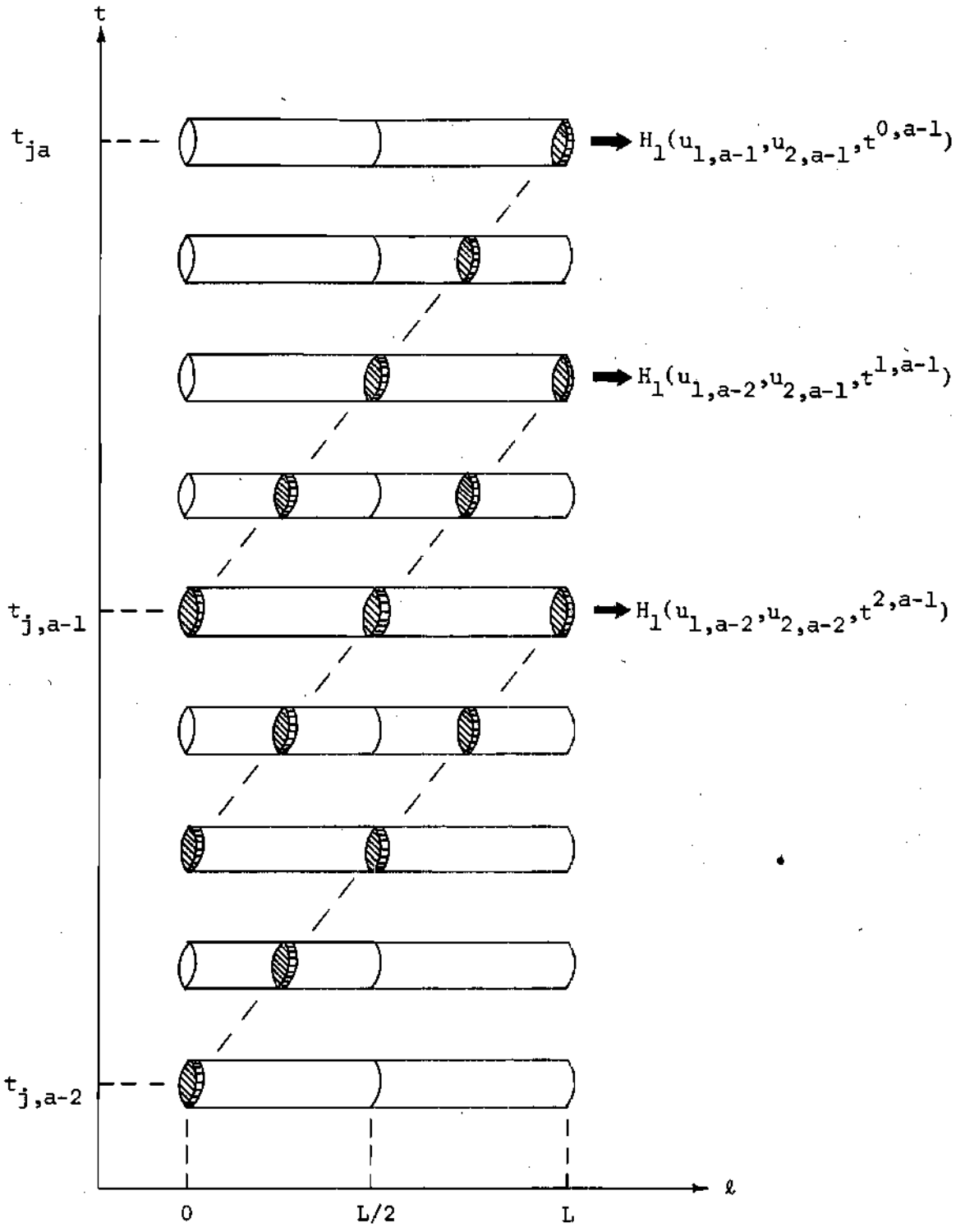


Figure 10. Progress of Successive Fluid Elements

$$\textcircled{2} : \quad H_1(u_{1,a-2}, u_{2,a-2}, t^{2,a-1})$$

$$\textcircled{1} : \quad H_1(u_{1,a-2}, u_{2,a-1}, t^{1,a-1})$$

$$\textcircled{0} : \quad H_1(u_{1,a-1}, u_{2,a-1}, t^{0,a-1})$$

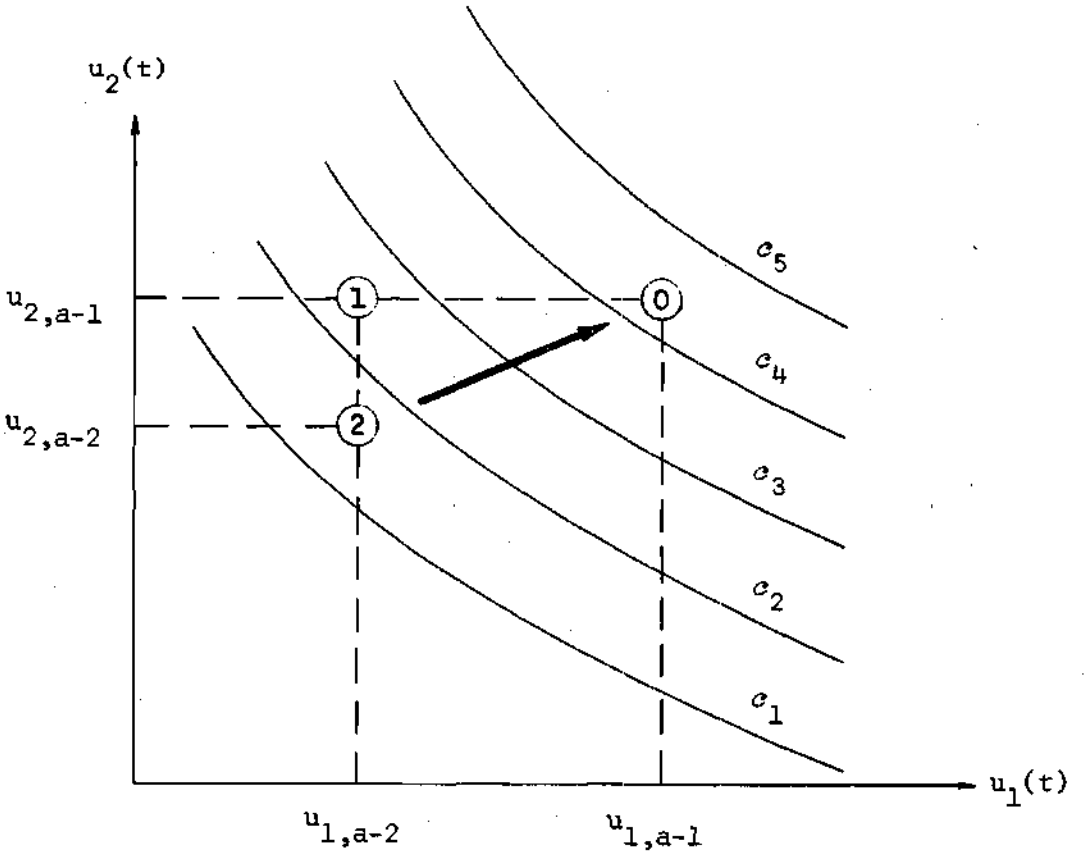


Figure 11. Contour Map of the Function $H_1(u_{1a}, u_{2a}, t^a)$

contour map of the function $H_1(u_{1a},u_{2a},t^a)$. Contours of constant $H_1(u_{1a},u_{2a},t^a)$ are denoted by c_i, where $c_{i+1} > c_i$. The three measurements plotted are seen to be linearly independent and, thus, to define a planar approximation to the $H_1(u_{1a},u_{2a},t^a)$ surface. Given this approximation, it is a simple matter to establish new set points which increase the value of $H_1(u_{1a},u_{2a},t^a)$.

The example of Figures 10 and 11 is undoubtedly an oversimplification. However, the basic technique is applicable to all of the systems modeled by (3-50) and (3-51). Development of this technique is continued in the next section.

Sequencing of the Algorithm

Consider now the equations governing the times at which the set points are changed and the times at which the measurements of (4-14) are made.

Origin of the Equations

The origin of these equations is in the partial derivative approximations of (4-14). The measurements required for all n of these approximations are

$$\underline{H}(u_{1,a-1},\ldots,u_{n,a-1},t^{0,a-1}) \tag{4-17}$$

$$\underline{H}(u_{1,a-2},u_{2,a-1},\ldots,u_{n,a-1},t^{1,a-1})$$

$$\underline{H}(u_{1,a-2},u_{2,a-2},u_{3,a-1},\ldots,u_{n,a-1},t^{2,a-1})$$

$$\vdots$$

$$\underline{H}(u_{1,a-2},\ldots,u_{n-2,a-2},u_{n-1,a-1},u_{n,a-1},t^{n-2,a-1})$$

$$\underline{H}(u_{1,a-2},\ldots,u_{n-1,a-2},u_{n,a-1},t^{n-1,a-1})$$

$$\underline{H}(u_{1,a-2},\ldots,u_{n,a-2},t^{n,a-1}).$$

Associated with each of these measurements is a validity condition, obtained from (3-51), specifying the time interval in which the exit states $\underline{x}(L,t^{j,a-1})$ are at steady state and, further, are characteristic of the particular combination of set points required. These validity conditions are given in Appendix D, along with other inequalities arising from the measuring instrument constraint (2-21).

Algorithmic Sequencing Equations

The inequalities noted above may be reduced to equalities by requiring that each algorithmic cycle be concluded in the minimum possible time. This reduction, given in Appendix D, establishes the set point command times $t_{j,a-1}$ and the measurement sample times $t^{j,a-1}$. The equations specifying these command and sample times are evaluated sequentially, starting with the $t^{n,a-1}$ equation and ending with the $t^{1,a-1}$ equation. From Appendix D, then, for stage n,

$$t^{n,a-1} = \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots \qquad (4\text{-}18)$$

$$+ \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L),\ t^{1,a-2} + \tau_m\}$$

$$t_{n,a-1} = t^{n,a-1} + \tau_m. \qquad (4\text{-}19)$$

For stage n-1,

$$t_{n-1,a-1} = \text{Max}\{t_{n,a-1}, \tag{4-20}$$

$$t_{n,a-1} - \rho_{n-1*}(L^*_{n-1}) + [\rho^*_n(L) - \sigma_{n*}(L)],$$

$$t^{n,a-1} + \tau_m - \rho_{n-1*}(L^*_{n-1}) - \sigma_{n*}(L)\}$$

$$t^{n-1,a-1} = t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L). \tag{4-21}$$

For stage n-2,

$$t_{n-2,a-1} = \text{Max}\{t_{n,a-1}, \tag{4-22}$$

$$t_{n-1,a-1} - \rho_{n-2*}(L^*_{n-2}) + [\rho^*_{n-1}(L^*_{n-1}) - \sigma_{n-1*}(L^*_{n-1})]$$

$$+ [\sigma^*_n(L) - \sigma_{n*}(L)],$$

$$t^{n-1,a-1} + \tau_m - \rho_{n-2*}(L^*_{n-2}) - \sigma_{n-1*}(L^*_{n-1}) - \sigma_{n*}(L)\}$$

$$t^{n-2,a-1} = t_{n-2,a-1} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L). \tag{4-23}$$

$$\vdots$$

For stage 2,

$$t_{2,a-1} = \text{Max}\{t_{n,a-1},$$ (4-24)

$$t_{3,a-1} - \rho_{2*}(L_2^*) + [\rho_3^*(L_3^*)-\sigma_{3*}(L_3^*)] + [\sigma_4^*(L_4^*)-\sigma_{4*}(L_4^*)]$$

$$+ \ldots + [\sigma_{n-1}^*(L_{n-1}^*)-\sigma_{n-1*}(L_{n-1}^*)] + [\sigma_n^*(L)-\sigma_{n*}(L)],$$

$$t^{3,a-1} + \tau_m - \rho_{2*}(L_2^*) - \sigma_{3*}(L_3^*) - \ldots - \sigma_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L)\}$$

$$t^{2,a-1} = t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L). \quad (4-25)$$

And for stage 1,

$$t_{1,a-1} = \text{Max}\{t_{n,a-1},$$ (4-26)

$$t_{2,a-1} - \rho_{1*}(L_1^*) + [\rho_2^*(L_2^*)-\sigma_{2*}(L_2^*)] + [\sigma_3^*(L_3^*)-\sigma_{3*}(L_3^*)]$$

$$+ \ldots + [\sigma_{n-1}^*(L_{n-1}^*)-\sigma_{n-1*}(L_{n-1}^*)] + [\sigma_n^*(L)-\sigma_{n*}(L)],$$

$$t^{2,a-1} + \tau_m - \rho_{1*}(L_1^*)-\sigma_{2*}(L_2^*) - \ldots - \sigma_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L)\}$$

$$t^{1,a-1} = t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L). \quad (4-27)$$

Since the measurement $\underline{H}(u_{1,a-1},\ldots,u_{n,a-1},t^{0,a-1})$ is the same as the measurement $\underline{H}(u_{1,a-2},\ldots,u_{n,a-2},t^{n,a-1})$ of the next algorithmic cycle, the measurement sample time $t^{0,a-1}$ is given by

$$t^{0,a-1} = t^{na}. \tag{4-28}$$

Application of the Sequencing Equations

It should be emphasized that the sequencing equations given above are not evaluated by the process-control computer of Figure 4 (page 20), but are evaluated off-line, generally by hand, and the results presented to the process-control computer as fixed program parameters. These parameters establish the sequencing of the 2n events occurring in each algorithmic cycle, namely the n set point adjustments at times $t_{j,a-1}$ and the n system measurements at times $t^{j,a-1}$. Given in Figure 12 is a schematic diagram of the sequencing of a typical algorithmic cycle. Also indicated in that figure is the point in the cycle at which the measured data are used to establish the next collection of set points.

Systems with Dominant Transport Delays

For systems in which the fluid transport delays are the dominant source of system transients, the set point adjustments tend to be concentrated at the beginning of the algorithmic cycle (as in Figure 12) and the measurement sample times tend to be spread uniformly throughout the algorithmic cycle. In many systems, in fact, the set point adjustments occur simultaneously and the algorithm operates in a manner analogous to Figures 10 and 11.

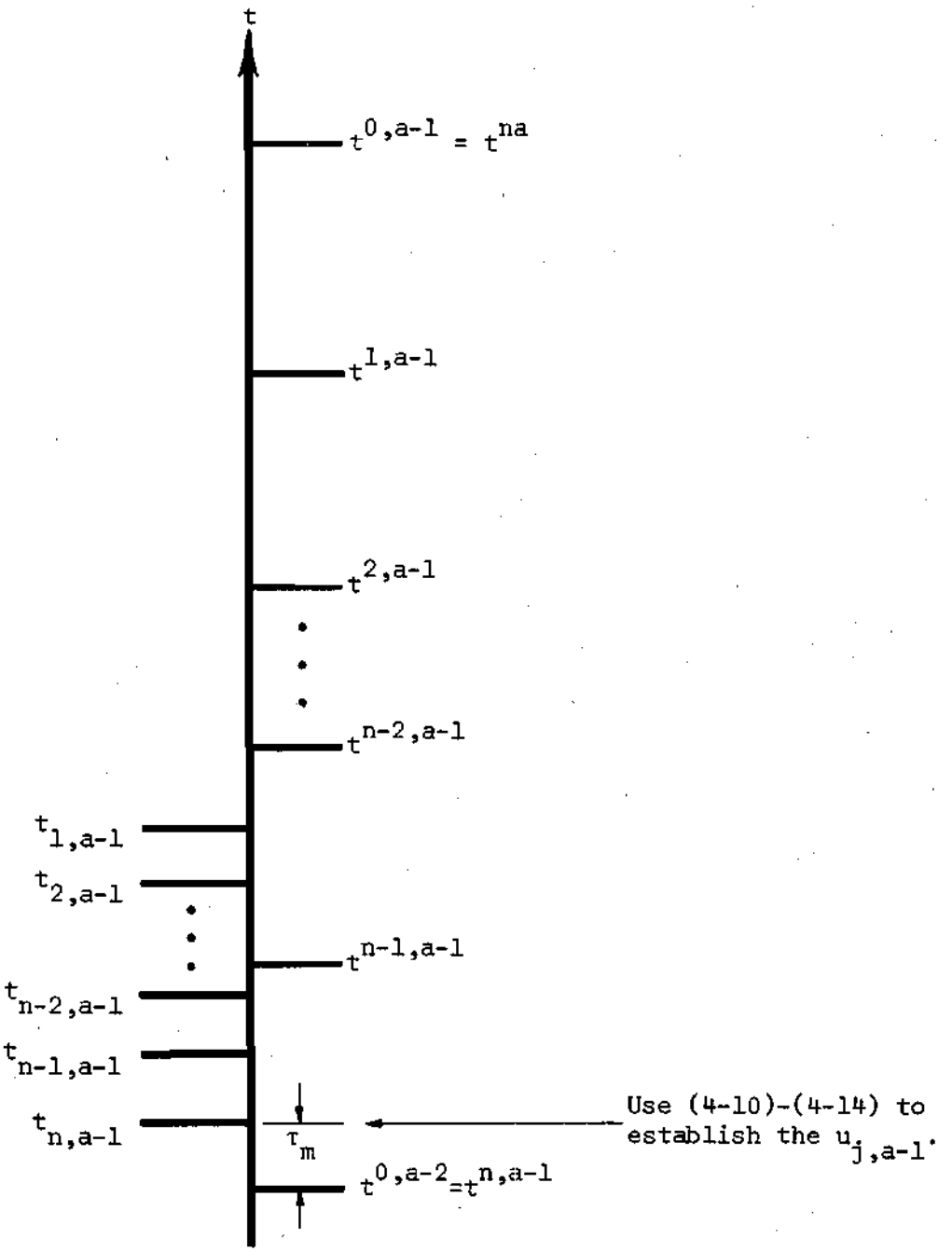


Figure 12. Sequencing of a Typical Algorithmic Cycle

Consider, for example, a two-stage system. Equations (4-18)-(4-21) are the applicable sequencing equations. Equation (4-20) reduces to

$$t_{n-1,a-1} = t_{n,a-1}, \tag{4-29}$$

implying that the set point adjustments occur simultaneously, if

$$t_{n,a-1} \geq t_{n,a-1} - \rho_{n-1*}(L^*_{n-1}) + [\rho^*_n(L) - \sigma_{n*}(L)] \tag{4-30}$$

and

$$t_{n,a-1} \geq t^{n,a-1} + \tau_m - \rho_{n-1*}(L^*_{n-1}) - \sigma_{n*}(L). \tag{4-31}$$

The inequality (4-31) is always satisfied, a conclusion which may be reached by inserting (4-19) into (4-31) and noting that $\rho_{n-1*}(L^*_{n-1})$ and $\sigma_{n*}(L)$ are both non-negative. The inequality (4-30) may be rewritten

$$\rho_{n-1*}(L^*_{n-1}) \geq \rho^*_n(L) - \sigma_{n*}(L). \tag{4-32}$$

This inequality is satisfied under rather broad conditions, best illustrated by using (B-10)-(B-13), basic stage form, to rewrite (4-32) as

$$\int_{L_{n-1}}^{L^*_{n-1}} \frac{1}{v^*(\ell)}\, d\ell \geq \tau_{nc} + \int_{L_{n*}}^{L} \frac{1}{v_*(\ell)}\, d\ell - \int_{L_{n*}}^{L} \frac{1}{v^*(\ell)}\, d\ell \tag{4-33}$$

$$= \tau_{nc} + \int_{L_{n*}}^{L} \frac{v^*(\ell) - v_*(\ell)}{v_*(\ell)v^*(\ell)}\, d\ell.$$

If the secondary interval $L_{n-1} < \ell \leq L^*_{n-1}$ is long enough, (4-33) is satisfied even with a significant controller response time τ_{nc} and a significant difference $v^*(\ell) - v_*(\ell)$ in the bounds on the fluid velocity. Thus, the multiplexed gradient sequencing equations are capable of compensating for the degrading effects of controller response time and fluid velocity uncertainty by making judicious use of the transport delay between sections to which control is applied. This compensation is the reason for the careful attention paid to the stage sub-intervals $L_{j*} < \ell \leq L_j$ and $L_j < \ell \leq L^*_j$ in the model development of Chapter III.

Operation of the Algorithm

Consider now the start-up, convergence, and closed-loop stability of the algorithm.

Start-Up

Let the time be $t^{[0]}$, and let the system be at steady state with the controls $u_j(t)$ maintained at set points $u_j^{[0]}$. Set "a" to 1 and define the u_{j0} by

$$u_{j0} = u_{j,a-1}\Big|_{a=1} = u_j^{[0]} \qquad j=1,\ldots,n. \tag{4-34}$$

Also define the measurement sample time t^{00} by

$$t^{00} = t^{0,a-1}\Big|_{a=1} = t^{[0]}. \tag{4-35}$$

At $t = t^{[0]}$, make the first system measurement

$$\underline{H}(u_{1,a-1},\ldots,u_{n,a-1},t^{0,a-1})\Big|_{a=1} = \underline{H}(u_{10},\ldots,u_{n0},t^{00}). \qquad (4\text{-}36)$$

At $t = t^{[0]} + \tau_m$, make an arbitrary choice for the $e_{ja}\big|_{a=1}$, say

$$e_{j1} = e_{ja}\big|_{a=1} = +1. \qquad (4\text{-}37)$$

The algorithm is then started, and its operation is governed by the equations previously given. The new set points, for example, are given by (4-10):

$$u_{j1} = u_{j0} + e_{j1}\Delta u_j. \qquad (4\text{-}38)$$

The measurement sample time t^{n1} is given by (4-28),

$$t^{n1} = t^{00}, \qquad (4\text{-}39)$$

and the stage n set point command time, after incrementing "a" to 2, by (4-19):

$$t_{n1} = t^{n1} + \tau_m. \qquad (4\text{-}40)$$

Convergence

Under the assumption of a linear relation between the u_{ja} and the function $F(u_{1a},\ldots,u_{na},t^a)$, and with no constraints on the system, the optimum is undefined. Proper algorithmic operation consists of increasing the value of $F(u_{1a},\ldots,u_{na},t^a)$ at each algorithmic cycle. This is easily shown to occur, for from (4-8),

$$F(u_{1a},\ldots,u_{na},t^a) - F(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}) \qquad (4\text{-}41)$$

$$= \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} (u_{ja}-u_{j,a-1}),$$

and from (4-10) and (4-11),

$$u_{ja} - u_{j,a-1} = \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j. \qquad (4\text{-}42)$$

It follows from (4-41) and (4-42) that

$$F(u_{1a},\ldots,u_{na},t^a) - F(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}) \qquad (4\text{-}43)$$

$$= \sum_{j=1}^{n} \left|\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right| \Delta u_j \geq 0.$$

Under the more general assumption of a concave unconstrained relation between the u_{ja} and the function $F(u_{1a},\ldots,u_{na},t^a)$, the optimum *is* defined and the algorithm can be shown to converge to a limit cycle about the optimum.[27]

Closed-Loop Stability

Consider finally the closed-loop stability of the algorithm. It has previously been required that the individual set point control loops be stable. It will now be shown that closing the algorithmic loop around the multistage system does not adversely affect system stability.

In this regard, suppose that the set points $u_{j,a-2}$ have been applied to the system. The last point in the system to be free of transients due to the application of these set points is the system exit $\ell = L$. This point is free of such transients by the time $t = t^{n,a-1}$ since the measurement $\underline{H}(u_{1,a-2},\ldots,u_{n,a-2},t^{n,a-1})$ is made only after the exit point has reached steady state.

Now consider the application of the next collection of set points, the $u_{j,a-1}$. An inspection of (4-18)-(4-27) reveals that the earliest time at which any of the $u_{j,a-1}$ are applied to the system is $t_{n,a-1}$. It follows that the entire system is free of transients due to set point changes in the time interval

$$[t^{n,a-1}, t_{n,a-1}] \tag{4-44}$$

which, from (4-19), is of duration $\tau_m \geq 0$. In view of this result, it may be concluded that the closed-loop system can support no sustained oscillations.

CHAPTER V

DEVELOPMENT OF THE CONSTRAINED ALGORITHM

The techniques developed in the previous chapter were extended to handle the control constraints of (2-15) and the state constraints of (2-18). Control constraints were handled by making a minor modification in the unconstrained algorithm. State constraints were handled by developing locally-linear models of the constrained variables and using the projected gradient technique[43,44] to establish new set points.

Constraints on the Controls

Consider first the control constraints of (2-15):

$$u_{j*} \leq u_j(t) \leq u_j^* \qquad j=1,\ldots,n. \tag{5-1}$$

Assuming that

$$u_j^* - u_{j*} \geq 2\Delta u_j, \tag{5-2}$$

(5-1) can be handled by replacing (4-10) with

$$u_{ja} = \begin{cases} u_{j,a-1} + e_{ja}\Delta u_j & u_{j*} \leq u_{j,a-1} + e_{ja}\Delta u_j \leq u_j^* \\ u_{j,a-1} - e_{ja}\Delta u_j & \text{otherwise.} \end{cases} \tag{5-3}$$

The operation of (5-3) is illustrated in Figure 13, where the set points resulting from several algorithmic cycles are shown on a contour map of the function $F(u_{1a},u_{2a},t^a)$. Successive set points are seen to take a path which zigzags along the constraint boundary, with the value of $F(u_{1a},u_{2a},t^a)$ increasing at alternate algorithmic cycles. The operation of the control-constrained algorithm is considered in more detail in Appendix E, where a constructive proof of its convergence is given.

Constraints on the States

Consider now the state constraints of (2-18):

$$z_{jkr*} \le z_{jkr}[\underline{x}(\ell_{jk},t)] \le z^*_{jkr} \tag{5-4}$$

$$j=1,\ldots,n \qquad k=1,\ldots,p_j \qquad r=1,\ldots,q_{jk}.$$

It was assumed that the transients at ℓ_{jk} were small enough that (5-4) could be replaced by the less restrictive condition

$$z_{jkr*} \le z_{jkr}[\underline{x}(\ell_{jk},t^{(jk)a})] \le z^*_{jkr} \tag{5-5}$$

$$j=1,\ldots,n \qquad k=1,\ldots,p_j \qquad r=1,\ldots,q_{jk}$$

where the $t^{(jk)a}$ were restricted to those portions of the algorithmic cycle in which the states at ℓ_{jk} were at steady state. The bounds on $t^{(jk)a}$ were obtained from (3-51) by setting ℓ to ℓ_{jk} and t to $t^{(jk)a}$.

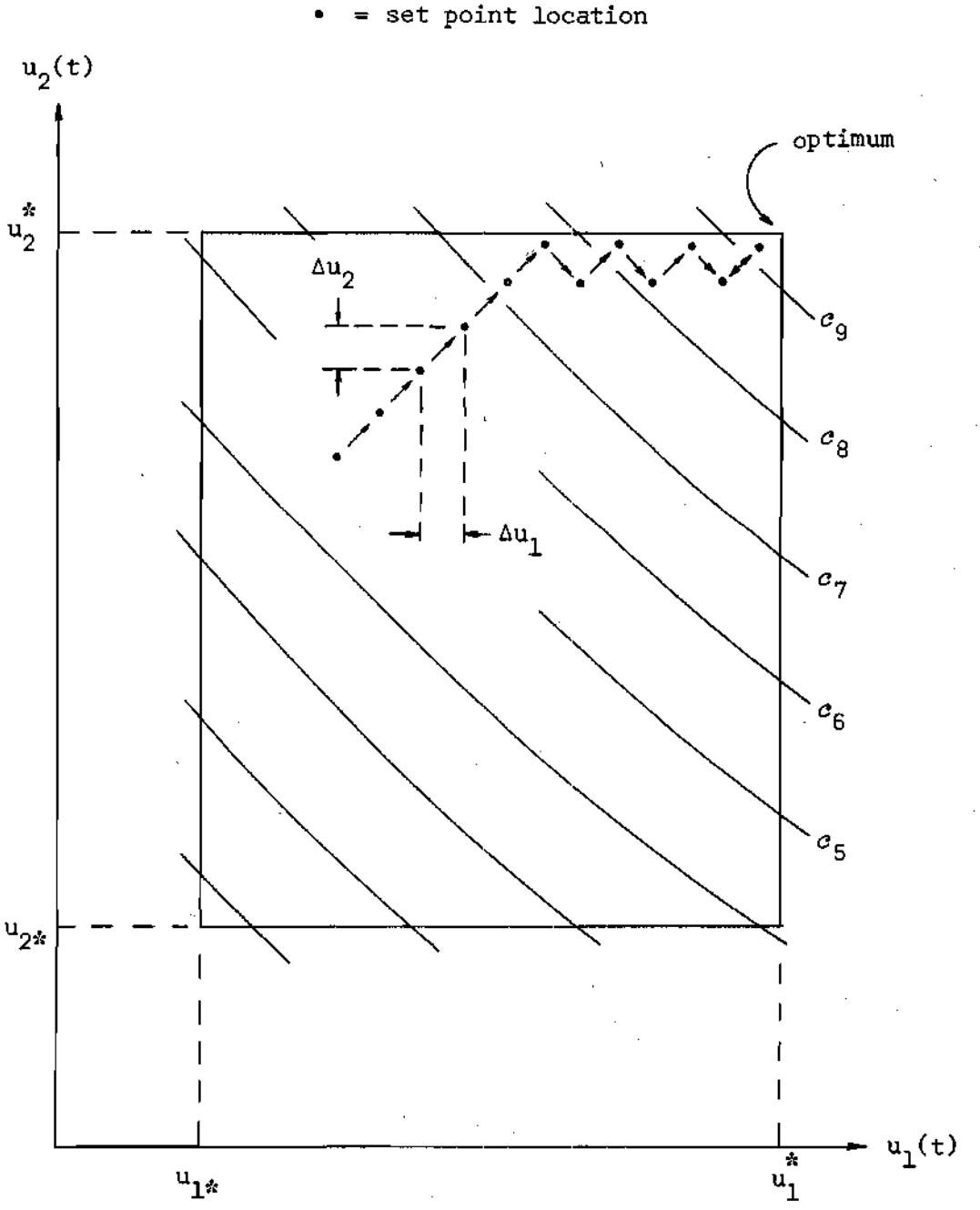


Figure 13. Contour Map of the Function $F(u_{1a},u_{2a},t^a)$

Locally-Linear Constraint Models

Locally-linear models of the constrained $z_{jkr}[\underline{x}(\ell_{jk},t^{(jk)a})]$ were established in a manner analogous to that used for the exit function $\underline{h}[\underline{x}(L,t^{a})]$. Specifically, the $\underline{x}(\ell_{jk},t^{(jk)a})$ were modeled by

$$\underline{x}(\ell_{jk},t^{(jk)a}) = \underline{Y}_j(\ell_{jk};u_{1a},\ldots,u_{ja},t^{(jk)a}) \tag{5-6}$$

and the functions Z_{jkr} were defined by

$$Z_{jkr}(u_{1a},\ldots,u_{ja},t^{(jk)a}) = z_{jkr}[\underline{x}(\ell_{jk},t^{(jk)a})]. \tag{5-7}$$

Then, the Z_{jkr} were expanded in Taylor series truncated at the linear terms,

$$Z_{jkr}(u_{1a},\ldots,u_{ja},t^{(jk)a}) = Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) \tag{5-8}$$

$$+ \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} (u_{ba}-u_{b,a-1}) + \left.\frac{\partial Z_{jkr}}{\partial t^{(jk)a}}\right|_{a-1} (t^{(jk)a}-t^{(jk)a-1}),$$

the assumption of small temporal drift was made,

$$\left|\left.\frac{\partial Z_{jkr}}{\partial t^{(jk)a}}\right|_{a-1} (t^{(jk)a}-t^{(jk)a-1})\right| << \left|\left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} (u_{ba}-u_{b,a-1})\right| \tag{5-9}$$

$$b=1,\ldots,j,$$

and (5-8) was simplified to

$$Z_{jkr}(u_{1a},\ldots,u_{ja},t^{(jk)a}) = Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) \qquad (5\text{-}10)$$

$$+ \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} (u_{ba}-u_{b,a-1}).$$

Finally, the partial derivatives of (5-10) were approximated by

$$\left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} = \frac{1}{(u_{b,a-1}-u_{b,a-2})} \times \qquad (5\text{-}11)$$

$$[Z_{jkr}(u_{1,a-2},\ldots,u_{b-1,a-2},u_{b,a-1},u_{b+1,a-1},\ldots,u_{j,a-1},t^{(jk)b-1,a-1})$$

$$- Z_{jkr}(u_{1,a-2},\ldots,u_{b-1,a-2},u_{b,a-2},u_{b+1,a-1},\ldots,u_{j,a-1},t^{(jk)b,a-1})].$$

Sequencing the Constrained Algorithm

Given in Appendix F is a derivation of the equations which establish, in the state-constrained case, the set point command times $t_{j,a-1}$, the exit measurement sample times $t^{j,a-1}$, and the constraint measurement times $t^{(jk)b,a-1}$. The equations governing the $t_{j,a-1}$ and $t^{j,a-1}$ are identical to those previously given as (4-18)-(4-28) and are not repeated here. The equations governing the $t^{(jk)b,a-1}$ are

$$t^{(jk)j,a-1} = t_{n,a-1} \qquad (5\text{-}12)$$

$$t^{(jk)j-1,a-1} = t_{j-1,a-1} + \rho_{j-1*}(L^{*}_{j-1}) + \sigma_{j*}(\ell_{jk})$$

$$t^{(jk)j-2,a-1} = t_{j-2,a-1} + \rho_{j-2*}(L^{*}_{j-2}) + \sigma_{j-1*}(L^{*}_{j-1}) + \sigma_{j*}(\ell_{jk})$$

$$\vdots$$

$$t^{(jk)2,a-1} = t_{2,a-1} + \rho_{2*}(L^{*}_{2}) + \sigma_{3*}(L^{*}_{3}) + \ldots + \sigma_{j-1*}(L^{*}_{j-1}) + \sigma_{j*}(\ell_{jk})$$

$$t^{(jk)1,a-1} = t_{1,a-1} + \rho_{1*}(L^{*}_{1}) + \sigma_{2*}(L^{*}_{2}) + \ldots + \sigma_{j-1*}(L^{*}_{j-1}) + \sigma_{j*}(\ell_{jk})$$

$$t^{(jk)0,a-1} = t^{(jk)a-1} = t^{(jk)ja}.$$

Method of Gradient Projection

Central to the state-constrained algorithm is the method of gradient projection.[43,44] In this regard, consider the gradient vector defined by

$$\underline{g} = \begin{bmatrix} \left.\frac{\partial F}{\partial u_{1a}}\right|_{a-1} \\ \vdots \\ \left.\frac{\partial F}{\partial u_{na}}\right|_{a-1} \end{bmatrix} \tag{5-13}$$

and the coordinate vector defined by

$$\underline{u}_a = \begin{bmatrix} u_{1a} \\ \vdots \\ u_{na} \end{bmatrix}. \tag{5-14}$$

Then consider the set of k linear constraints

$$\underline{y}_i^T \underline{u}_a - b_i \geq 0 \qquad i=1,\ldots,k, \tag{5-15}$$

where the superscript T denotes a vector transpose. Given in Appendix G are definitions which permit the constraints of (5-5) to be written in the form of (5-15).

Consider now the point $\underline{u}_{a-1}$, and let it be in the intersection of a subset of the k constraints considered in (5-15):

$$\underline{y}_i^T \underline{u}_{a-1} - b_i = 0 \qquad i=1,\ldots,q\leq k. \tag{5-16}$$

A procedure has been developed[43] for extracting a set of q linearly independent constraints from this set of q constraints, and for generating an n×n matrix M_q with the property that the vector $\underline{w}$ given by

$$\underline{w} = M_q \underline{g} \tag{5-17}$$

is parallel to the intersection of the q constraints and is in the direction of increasing $F(u_{1a},\ldots,u_{na},t^a)$. The vector $\underline{w}$ is called the projected gradient.

The vectors $\underline{g}$, $\underline{y}_i$, and $\underline{w}$ are illustrated in Figure 14 for the case n=2, k=4, and q=q=1. The boundaries of Figure 14 enclose the collection of points which satisfy the constraints of (5-15).

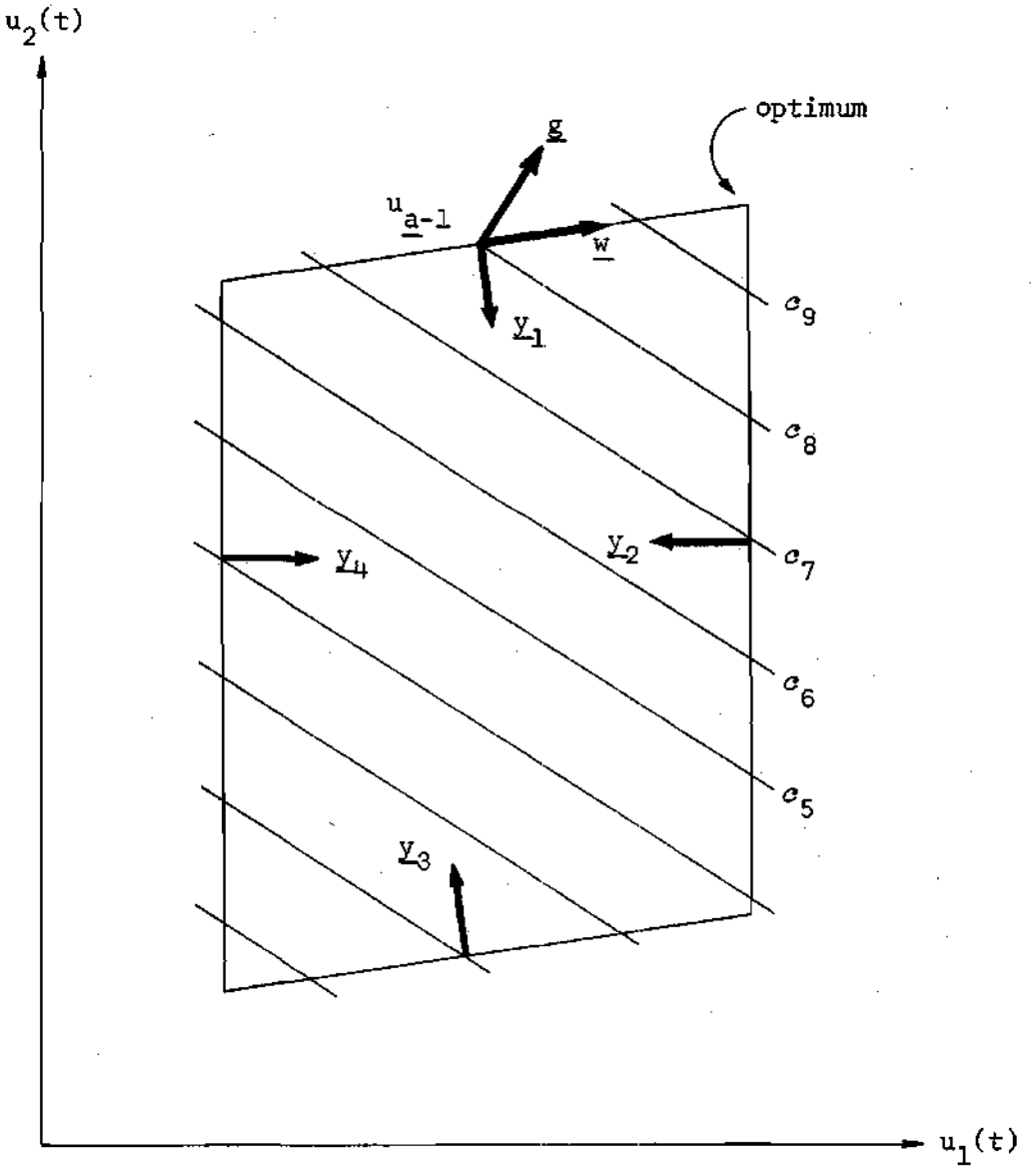


Figure 14. Linearized Contour Map of $F(u_{1a},u_{2a},t^a)$

The procedure[43] for generating the matrix M_q is given in Appendix G. Also given is an alternative procedure, developed as a part of the thesis research, which eliminates the need for performing division and square root operations.

Tracking the Projected Gradient

Consider now a procedure for tracking the projected gradient with a series of set point adjustments of magnitude Δu_j. First, let the vector $\underline{w}$ have components w_j:

$$\underline{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}. \tag{5-18}$$

Next, consider the ordered n-tuple $\underline{\psi} = (\psi_1,\ldots,\psi_n)$ whose elements are the reordered elements of the n-tuple $(1,\ldots,n)$. The elements of $\underline{\psi}$ are ordered so that

$$\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} \geq \frac{|w_{\psi_2}|}{\Delta u_{\psi_2}} \geq \cdots \geq \frac{|w_{\psi_n}|}{\Delta u_{\psi_n}}. \tag{5-19}$$

Then define the scalar SGN by

$$SGN = \text{sign}\left[\sum_{k=2}^{n} \frac{|w_{\psi_k}|}{\Delta u_{\psi_k}} - (n-3)\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}}\right] \tag{5-20}$$

and the n vectors

$$\underline{c}_k = \begin{bmatrix} c_{k1} \\ \vdots \\ c_{kn} \end{bmatrix} \qquad k=1,\ldots,n \tag{5-21}$$

by

$$c_{kj} = \begin{cases} \text{SGN sign}(w_j)\Delta u_j & k=1 \quad j=1,\ldots,n \\ + \ \text{sign}(w_j)\Delta u_j & k=2,\ldots,n \quad j=1,\ldots,n \quad \psi_k \neq j \\ - \ \text{sign}(w_j)\Delta u_j & k=2,\ldots,n \quad j=1,\ldots,n \quad \psi_k = j. \end{cases} \tag{5-22}$$

The $\underline{c}_k$ are linearly independent and, thus, span the n-dimensional Euclidean space in which the vector $\underline{w}$ lies. It follows that there is a unique linear combination of the $\underline{c}_k$ which equals the vector $\underline{w}$:[45]

$$d_1\underline{c}_1 + d_2\underline{c}_2 + \ldots + d_n\underline{c}_n = \underline{w}. \tag{5-23}$$

It is shown in Appendix H that the d_k of (5-23) are the non-negative quantities

$$d_1 = 0.5 \ \text{SGN}\left[\sum_{k=2}^{n} \frac{|w_{\psi_k}|}{\Delta u_{\psi_k}} - (n-3)\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}}\right] \tag{5-24}$$

$$d_2 = 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_2}|}{\Delta u_{\psi_2}}\right]$$

$$\vdots$$

$$d_n = 0.5\left[\frac{|^w\psi_1|}{\Delta u_{\psi_1}} - \frac{|^w\psi_n|}{\Delta u_{\psi_n}}\right].$$

Now let (5-23) be multiplied by a positive constant K whose magnitude is such that all of the products $Kd_1,\ldots,Kd_n$ are integers:

$$Kd_1\underline{c}_1 + Kd_2\underline{c}_2 + \ldots + Kd_n\underline{c}_n = K\underline{w}. \qquad (5\text{-}25)$$

Then let β^* be the sum of the Kd_k:

$$\beta^* = \sum_{k=1}^{n} Kd_k. \qquad (5\text{-}26)$$

Given in Appendix H is a procedure for establishing a β^*-tuple $\underline{\pi} = (\pi_0,\ldots,\pi_{\beta^*-1})$ whose elements are the reordered elements of the β^*-tuple $(\underbrace{1,\ldots,1}_{Kd_1},\ldots,\underbrace{n,\ldots,n}_{Kd_n})$. The elements of $\underline{\pi}$ are ordered so that the collection of points

$$\underline{u}_{a+\beta} = \underline{u}_{a-1} + \sum_{b=0}^{\beta} \underline{c}_{\pi_b} \qquad \beta=0,\ldots,\beta^*-1 \qquad (5\text{-}27)$$

tracks the projected gradient. Each $\underline{c}_{\pi_b}$ is one of the $\underline{c}_k$ of (5-25). The subscript π_b is one of the set $\pi_0,\ldots,\pi_{\beta^*-1}$, and specifies which $\underline{c}_k$ is used in the sum of (5-27). Using (5-25), it may be seen that the $\underline{u}_{a+\beta}$ defined by $\beta=\beta^*-1$ lies directly on the projected gradient:

$$\underline{u}_{a+\beta^*-1} = \underline{u}_{a-1} + \sum_{b=0}^{\beta^*-1} \underline{c}_{\pi_b} \tag{5-28}$$

$$= \underline{u}_{a-1} + \sum_{k=1}^{n} Kd_k \underline{c}_k$$

$$= \underline{u}_{a-1} + K\underline{w}.$$

The tracking procedure outlined above is illustrated in Figure 15 for the case n = 2, $\underline{w} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, and $\Delta u_1 = \Delta u_2 = 2$. For this case, (5-19)-(5-22) give $\underline{c}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\underline{c}_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$, and (5-24) gives $d_1 = 1.5$ and $d_2 = 1.0$. For the products Kd_1 and Kd_2 to both be integers, the multiplier K has to be 2. Equation (5-26) then gives $\beta^* = 5$, the 5-tuple $\underline{\pi}$ is (1,2,1,2,1), and the $\underline{u}_{a+\beta}$ of (5-27) are

$$\underline{u}_a = \underline{u}_{a-1} + \underline{c}_1 \tag{5-29}$$

$$\underline{u}_{a+1} = \underline{u}_{a-1} + \underline{c}_1 + \underline{c}_2$$

$$\underline{u}_{a+2} = \underline{u}_{a-1} + \underline{c}_1 + \underline{c}_2 + \underline{c}_1$$

$$\underline{u}_{a+3} = \underline{u}_{a-1} + \underline{c}_1 + \underline{c}_2 + \underline{c}_1 + \underline{c}_2$$

$$\underline{u}_{a+4} = \underline{u}_{a-1} + \underline{c}_1 + \underline{c}_2 + \underline{c}_1 + \underline{c}_2 + \underline{c}_1.$$

These points, shown in Figure 15, track the projected gradient $\underline{w}$.

Aligning the Tracking Cycle

Consider now a generalization of the tracking procedure outlined above. First, let the vector $\underline{w}$ be computed at each algorithmic cycle.

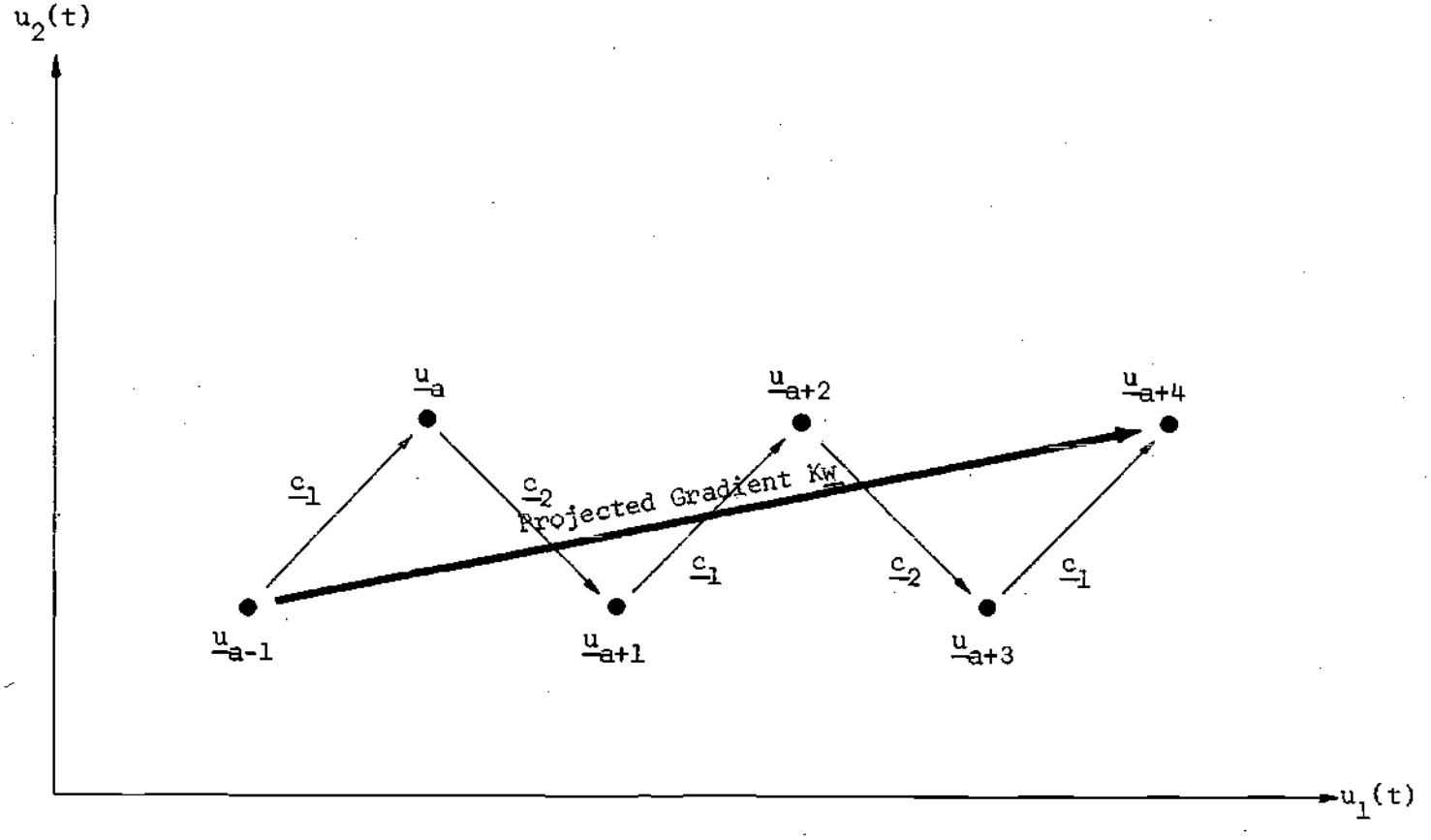


Figure 15. Tracking the Projected Gradient

Then, extend the β^*-tuple $\underline{\pi}$ to a $2\beta^*$-tuple

$$\underline{\pi} = (\pi_0,\ldots,\pi_{\beta^*-1},\pi_{\beta^*},\ldots,\pi_{2\beta^*-1}) \tag{5-30}$$

by repeating the elements of the original $\underline{\pi}$:

$$\pi_\beta = \pi_{\beta+\beta^*} \qquad \beta=0,\ldots,\beta^*-1. \tag{5-31}$$

Next, consider the collection of set points

$$\underline{u}_{a+\beta,\beta'} = \underline{u}_{a-1} + \sum_{b=0}^{\beta} \underline{c}_{\pi_{b+\beta'}} \qquad \begin{matrix}\beta=0,\ldots,\beta^*-1 \\ \beta'=0,\ldots,\beta^*-1.\end{matrix} \tag{5-32}$$

For each β', the collection of points $\underline{u}_{a+\beta,\beta'}$ is called a tracking cycle, with the integer β' specifying the cycle alignment. Each tracking cycle contains the same collection of $\underline{c}_k$ as employed in (5-27), with β' simply specifying the cycle starting point.

Now let $E'_{\beta i}$ be the distance from the ith constraint boundary (i=1,...,q: see 5-17) to that point in the β'th tracking cycle which is closest to the ith constraint boundary:

$$E'_{\beta i} = \operatorname*{Min}_{\beta=0}^{\beta^*-1} \{\underline{y}_i^T \underline{u}_{a+\beta,\beta'} - b_i\}. \tag{5-33}$$

Then let β'' be that alignment β' which minimizes the sum

$$\sum_{i=1}^{q} E'_{\beta i} \qquad \beta'=0,\ldots,\beta^*-1 \tag{5-34}$$

subject to the constraints

$$E'_{\beta i} \geq 0 \qquad \begin{matrix} \beta=0,\ldots,\beta^{*}-1 \\ i=1,\ldots,q. \end{matrix} \tag{5-35}$$

The equation specifying the set point $\underline{u}_a$, replacing in the state-constrained case (4-10) and (4-11), is

$$\underline{u}_a = \underline{u}_{a-1} + \underline{c}_{\pi''_\beta}. \tag{5-36}$$

The basic philosophy of the procedure given by (5-32)-(5-36) is illustrated in Figure 16 for the case n=2 and q=1. It is helpful to think of the constraint boundary and the point $\underline{u}_{a-1}$ as being fixed and the collection of $\underline{c}_k$ tracking the projected gradient as being movable. The collection of $\underline{c}_k$ is then given various trial alignments. The alignment chosen is the one which provides the closest fit, without violation, to the constraint boundary. The point $\underline{u}_a$ is then taken as the next point in the tracking cycle after $\underline{u}_{a-1}$.

It is shown in Appendix H that the continued application of this procedure results in the selection of set points which continue to follow the tracking cycle. Since the tracking cycle follows the projected gradient, and the projected gradient converges on the constrained optimum,[43] it follows that the state-constrained algorithm converges to the constrained optimum. Further, since the algorithm uses only the latest locally-linear system model and requires no memory of the previous cycle

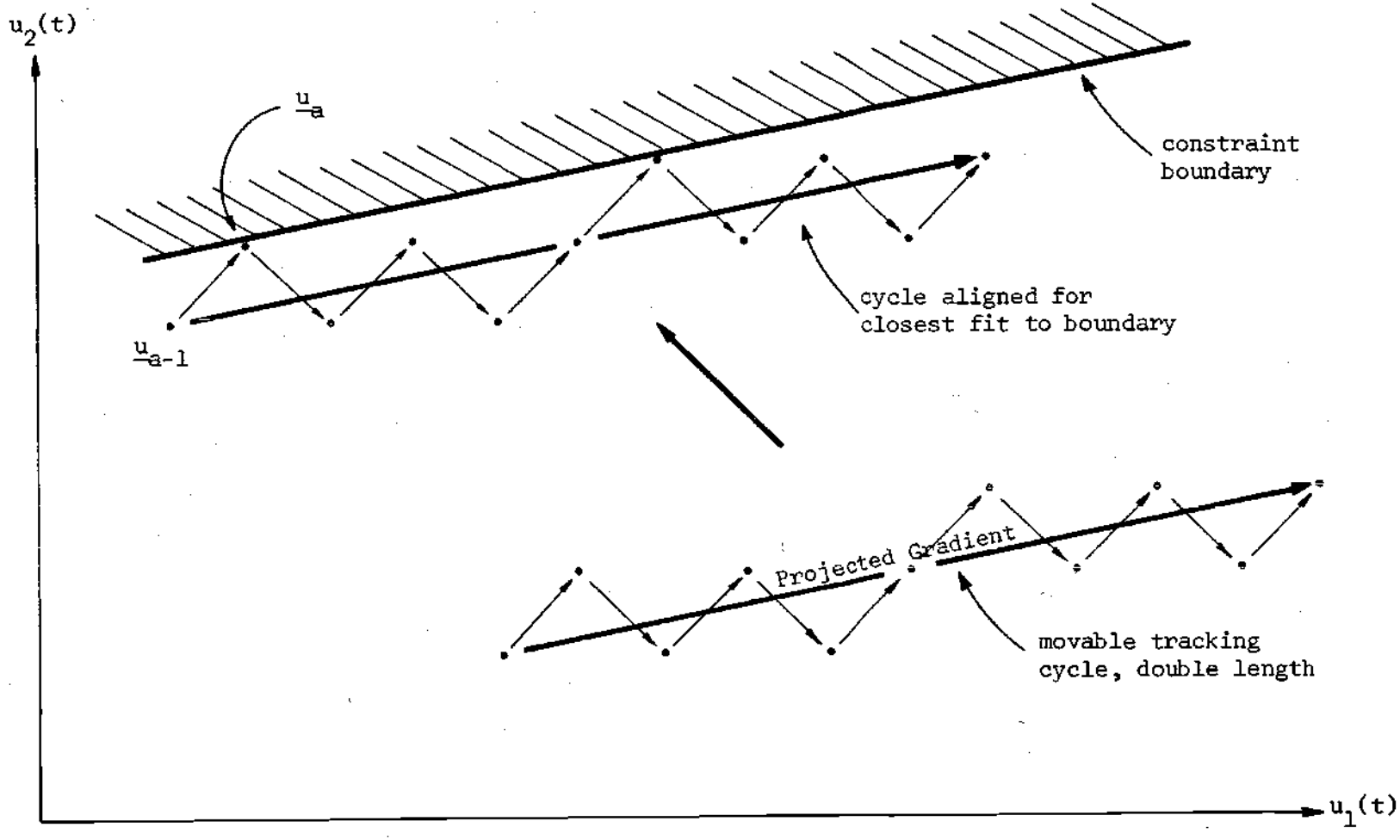


Figure 16. Aligning the Tracking Cycle

alignment, it can follow a time-varying projected gradient whether the variation is due to static system nonlinearities or dynamic system parameters.

Establishing the Constraint Set

The procedure outlined above results in set points which are close to, but not necessarily on, the constraint boundaries of (5-16). It is necessary, therefore, to replace (5-16) by the less restrictive condition

$$\underline{y}_i^T \underline{u}_{a-1} - b_i \leq \epsilon_i \qquad i=1,\ldots,q \tag{5-37}$$

where the ϵ_i are chosen to insure that (a) constraints not satisfying (5-37) are far enough from $\underline{u}_{a-1}$ to be safely ignored in the computation of $\underline{u}_a$ and (b) constraints included in the set i=1,...,q continue to satisfy (5-37) in subsequent algorithmic cycles. Condition (a) is to insure continued constraint satisfaction and condition (b) is to insure continued tracking of the projected gradient.

Given in Appendix I is an analysis which establishes an upper bound on the width of the tracking cycle. It is shown in that appendix that the choice

$$\epsilon_i = (n+2)|y_{iJ}|\Delta u_J + \left(\frac{n^2}{2} + \frac{n}{2} - 2\right) \sum_{j=1}^{n} |y_{ij}|\Delta u_j, \tag{5-38}$$

where J is that index j for which

$$|y_{iJ}|\Delta u_J \geq |y_{ij}|\Delta u_j \qquad j=1,\ldots,J-1,J+1,\ldots,n, \tag{5-39}$$

insures conditions (a) and (b). Equations (5-37)-(5-39) thus establish the set of constraints to be considered for inclusion in the projection matrix M_q.

Approximating the Tracking Cycle

If β^* should be large, say on the order of 100, the computational effort required to solve the minimization problem (5-33)-(5-35) would be massive. It is necessary, therefore, to establish some approximation to the basic tracking procedure which permits control over β^*. The approximation used was to round the largest d_k to ω significant binary places, and to round the other d_k at the same bit position. Denoting the rounded d_k by $d_k^{(r)}$, it follows that multiplication by the constant K may be accomplished by relocating the binary point at the right of the least significant bit of the largest $d_k^{(r)}$. It also follows that the products $Kd_k^{(r)}$ are all integers, that they satisfy the inequalities

$$Kd_k^{(r)} \leq 2^{\omega} \qquad k=1,\ldots,n, \tag{5-40}$$

and that the β^* resulting from the rounded d_k satisfies the inequality

$$\beta^{*(r)} = \sum_{k=1}^{n} Kd_k^{(r)} \leq n2^{\omega}. \tag{5-41}$$

Choice of the parameter ω is a compromise between precision of tracking (large ω) and speed of computation (small ω). Guidelines for this choice are presented in Chapter VIII.

Concluding Remarks

Several remarks are appropriate at this point. First, if no constraints are close enough to $\underline{u}_{a-1}$ to satisfy (5-37), or if the unconstrained move is directed away from those constraints which do satisfy (5-37), then the unconstrained algorithm should be used.

Second, since the tracking cycle using the rounded d_k is only approximately parallel to the q constraint boundaries, it may happen that there is no β' which satisfies all of the constraints of (5-35). If this occurs, let β'' be that alignment which maximizes

$$\min_{i=1}^{q}\{E'_{\beta i}\}. \tag{5-42}$$

This procedure will correct for impending constraint violations several algorithmic cycles before they would occur.

Finally, when q=n it may be concluded that $\underline{u}_{a-1}$ is at the optimum.[43] When this occurs, let the u_{ja} be given by (4-10) with the $e_{ja} = \pm 1$ chosen to maximize $F(u_{1a},\ldots,u_{na},t^a)$ subject to the constraints of (5-37). This procedure will establish, and maintain, a limit cycle about the optimum.

Constraints on the Controls and the States

Consider now the control of systems with both control and state constraints. The control constraints of (2-15), with $u_j(t) = u_{ja}$, may be written

$$\begin{aligned} u_{ja} - u_{j*} &\geq 0 \\ & \qquad j=1,\ldots,n. \\ -u_{ja} + u_j^* &\geq 0 \end{aligned} \tag{5-43}$$

These in turn may be put in the form of (5-15) by defining, for $j=1,\ldots,n$ and $i=k+j$,

$$y_{ib} = \begin{cases} 1 & b=j \\ 0 & b=1,\ldots,j-1,j+1,\ldots,n \end{cases} \tag{5-44}$$

and

$$b_i = u_{j*}, \tag{5-45}$$

and by defining, for $j=1,\ldots,n$ and $i=k+n+j$,

$$y_{ib} = \begin{cases} -1 & b=j \\ 0 & b=1,\ldots,j-1,j+1,\ldots,n \end{cases} \tag{5-46}$$

and

$$b_i = -u_j^*. \tag{5-47}$$

Thus, the control constraints may be treated as state constraints whose locally-linear models are known a priori, and systems with both control and state constraints may be handled by the state-constrained algorithm.

CHAPTER VI

ANALYSIS OF MEASUREMENT NOISE EFFECTS

The preceding chapters have outlined the development of the unconstrained, control-constrained, and state-constrained extremum control algorithms. The remaining chapters establish some of the properties of these algorithms. In this chapter the effects of additive white Gaussian measurement noise on the operation of the unconstrained algorithm are examined. Expressions for the mean and standard deviation of the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$ are given, and from these expressions the performance of the algorithm in a noise-corrupted environment is predicted.

Mean of Per-Cycle Increase

The computation of the per-cycle increase begins with the assumption that each of the measurements of (4-14) is corrupted by samples of a vector white Gaussian process with mean $\underline{\eta}$ and standard deviation $\underline{\sigma}$. Denoting noise-corrupted quantities by the symbol ^, it follows from (4-14) that

$$\left.\frac{\partial \hat{H}}{\partial u_{ja}}\right|_{a-1} = \frac{1}{(u_{j,a-1}-u_{j,a-2})} \times \tag{6-1}$$

$$[\underline{H}(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-1},u_{j+1,a-1},\ldots,u_{n,a-1},t^{j-1,a-1}) + \chi^{j-1,a-1}$$

$$- \underline{H}(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-2},u_{j+1,a-1},\ldots,u_{n,a-1},t^{j,a-1}) - \underline{\chi}^{j,a-1}]$$

where $\underline{\chi}^{j,a-1}$ represents the samples of the measurement noise process. It similarly follows from (4-13) that

$$\frac{\partial \hat{F}}{\partial u_{ja}} = \frac{\partial G}{\partial H_1}\frac{\partial \hat{H}_1}{\partial u_{ja}} + \ldots + \frac{\partial G}{\partial H_\mu}\frac{\partial \hat{H}_\mu}{\partial u_{ja}} + \frac{\partial G}{\partial u_{ja}} \tag{6-2}$$

and from (4-11) that

$$e_{ja} = \text{sign}\left[\frac{\partial \hat{F}}{\partial u_{ja}}\bigg|_{a-1}\right]. \tag{6-3}$$

Defining the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$ as

$$\Delta F = F(u_{1a},\ldots,u_{na},t^a) - F(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}), \tag{6-4}$$

it follows from (4-8), (4-10), and (6-3) that

$$\Delta F = \sum_{j=1}^{n} \frac{\partial F}{\partial u_{ja}}\bigg|_{a-1} \text{sign}\left[\frac{\partial \hat{F}}{\partial u_{ja}}\bigg|_{a-1}\right]\Delta u_j. \tag{6-5}$$

It is shown in Appendix J that the expected value of this increase, denoted by $E\{\Delta F\}$, is for $n \geq 2$ given by

$$E\{\Delta F\} = \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j 2\,\mathrm{erf}\left[\frac{\left(\sum_{i=1}^{\mu} \left.\frac{\partial G}{\partial H_i}\right|_{a-1} \left.\frac{\partial H_i}{\partial u_{ja}}\right|_{a-1} + \left.\frac{\partial G}{\partial u_{ja}}\right|_{a-1}\right)\Delta u_j}{\sqrt{\sum_{i=1}^{\mu}\left(\left.\frac{\partial G}{\partial H_i}\right|_{a-1} \sqrt{2}\,\sigma_i\right)^2}}\right] \tag{6-6}$$

where the function erf(x) is given by

$$\mathrm{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy. \tag{6-7}$$

For the frequently encountered case of maximizing a single system output, $\mu=1$ and $F(u_{1a},\ldots,u_{na},t^a) = H_1(u_{1a},\ldots,u_{na},t^a)$. In this case (6-6) reduces to

$$E\{\Delta F\} = \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j 2\,\mathrm{erf}\left[\frac{\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j}{\sqrt{2}\,\sigma_1}\right]. \tag{6-8}$$

The expression given in (6-8) is plotted in Figure 17 for the case

$$\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} = \left.\frac{\partial F}{\partial u_{1a}}\right|_{a-1} \qquad j=1,\ldots,n, \tag{6-9}$$

along with results from a Monte Carlo simulation for the case $n=1$. The ordinate of Figure 17 is scaled so that with no measurement noise (that is, with the noise standard deviation σ_1 equal to zero) the expected

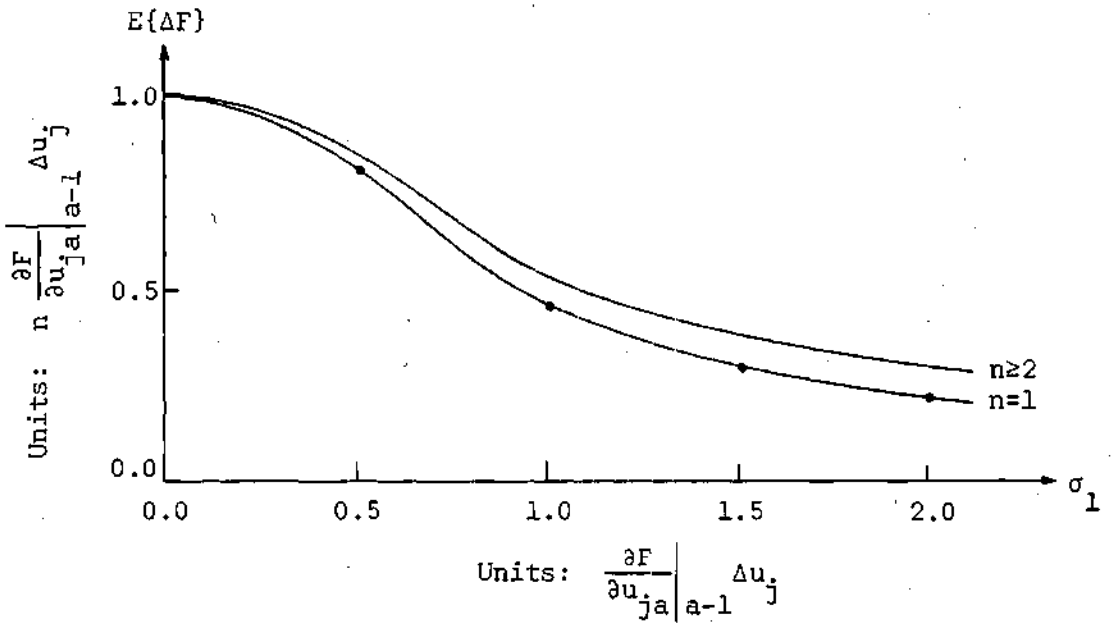


Figure 17. Mean of the Increase in $F(u_{1a},\ldots,u_{na},t^a)$

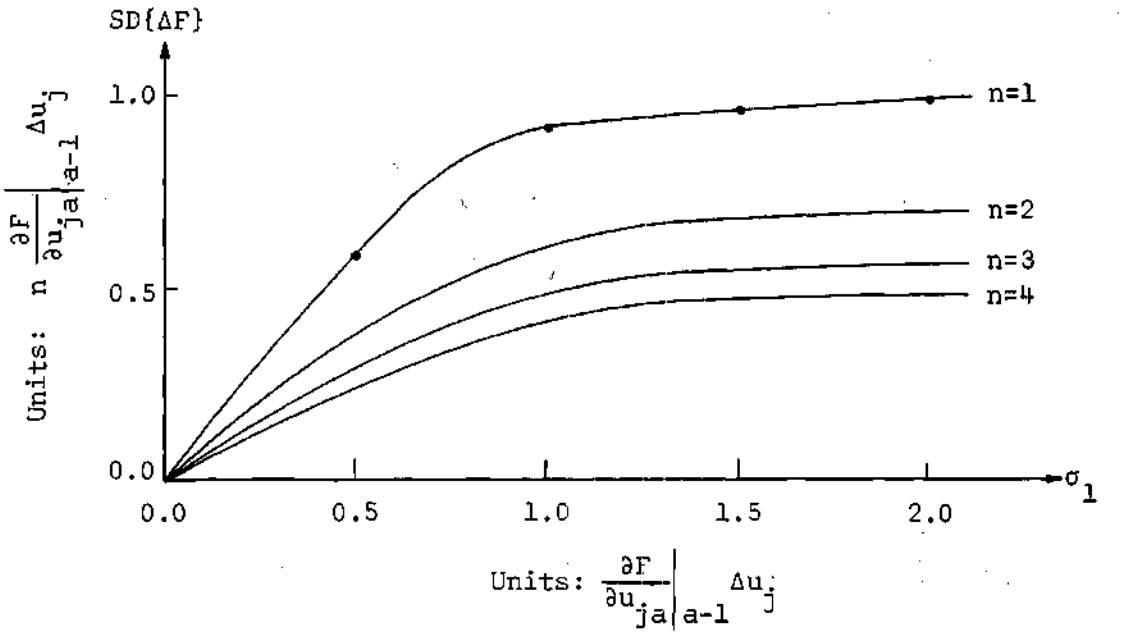


Figure 18. Standard Deviation of the Increase in $F(u_{1a},\ldots,u_{na},t^a)$

value of the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$ is 1.0. As the noise level increases (that is, as σ_1 increases), the expected value of ΔF decreases as shown. The abscissa of Figure 17 is scaled in units of

$$\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j, \tag{6-10}$$

which is the finite difference whose sign must be correctly determined if the algorithm is to adjust the set points in the proper direction. It may be concluded from Figure 17 that the algorithm is relatively insensitive to measurement noise as long as

$$\sigma_1 \leq 0.5 \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j. \tag{6-11}$$

With proper choice of the parameter Δu_j, (6-11) should be satisfied for all set points except those quite close to the optimum.

Standard Deviation of Per-Cycle Increase

An expression for the standard deviation of the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$, denoted by SD{$\Delta F$}, is given in Appendix J. This expression was evaluated numerically for the case considered in Figure 17, and the results of that evaluation are given in Figure 18. At low noise levels, the standard deviation may be seen to be a linear function of the ratio $\sigma_1 \Big/ \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j$. At high noise levels, it approaches a constant. An examination of (J-34) in the limit as $\sigma_1 \to \infty$ reveals that this constant varies inversely with the square root of the number of system stages.

For σ_1 satisfying (6-11), Figures 17 and 18 together yield the inequality

$$E\{\Delta F\} - SD\{\Delta F\} > 0. \qquad (6\text{-}12)$$

It follows from (6-12), somewhat loosely, that when (6-11) is satisfied the algorithm may be expected to increase the value of $F(u_{1a},\ldots,u_{na},t^a)$ at *each* algorithmic cycle.

CHAPTER VII

ANALYSIS OF HARDWARE REQUIREMENTS

The hardware required to implement the unconstrained, control-constrained, and state-constrained algorithms was compared to that required to implement two currently available procedures. Storage and word length requirements were established, and computation times were estimated.

Currently Available Procedures

The first of the two procedures considered was the one-at-a-time technique,[22] a method frequently used because of its computational simplicity. This procedure applies (4-10)-(4-13) to one set point at a time, approximating the partial derivatives of (4-13) by

$$\left.\frac{\partial \underline{H}}{u_{ja}}\right|_{a-1} = \frac{1}{(u_{j,a-1}-u_{j,a-2})} \times \tag{7-1}$$

$$[\underline{H}(u_{10},\ldots,u_{j-1,0},u_{j,a-1},u_{j+1,0},\ldots,u_{n0},t^{j,a-1})$$

$$- \underline{H}(u_{10},\ldots,u_{j-1,0},u_{j,a-2},u_{j+1,0},\ldots,u_{n0},t^{j,a-2})].$$

The first adjustment of u_{ja} is arbitrary and exploratory. Subsequent adjustments based on (4-10)-(4-13) and (7-1) are continued until e_{ja} reverses sign or a constraint boundary is encountered.

The second procedure considered was the standard gradient technique[25] operating in conjunction with the projected gradient technique,[43,44] a combination which has been quite effective in constrained nonlinear programming problems.[47] This procedure makes n exploratory set point adjustments to establish the n partial derivative approximations

$$\left.\frac{\partial \underline{H}}{\partial u_{ja}}\right|_{a-1} = \frac{1}{\Delta u_j}\left[\underline{H}(u_{10},\ldots,u_{j-1,0},u_{j0}+\Delta u_j,u_{j+1,0},\ldots,u_{n0},t^j)\right. \qquad (7\text{-}2)$$

$$\left. - \underline{H}(u_{10},\ldots,u_{n0},t^0)\right],$$

and then adjusts all of the set points simultaneously by an amount proportional to the gradient $\underline{g}$ or the projected gradient $\underline{w}$. Adjustments proportional to $\underline{g}$ or $\underline{w}$ are continued until $F(u_{1a},\ldots,u_{na},t^a)$ decreases or a new constraint boundary is encountered.

Storage and Computation Time Requirements

Consider now the storage and computation time requirements of the unconstrained multiplexed gradient technique. Implementation of (4-18)-(4-27) requires storing 2n system parameters (to define the times, relative to $t^{n,a-1}$, of the 2n events shown in Figure 12, page 52), performing 2n additions per algorithmic cycle (to convert these relative times to real times), and storing 2n results (the converted real times). Equations (4-11)-(4-14), restricting the analysis to the frequently encountered case of μ=1 and $F(u_{1a},\ldots,u_{na},t^a) = H_1(u_{1a},\ldots,u_{na},t^a)$, may be simplified to

$$e_{ja} = e_{j,a-1} \times \tag{7-3}$$

$$\text{sign}[F(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-1},u_{j+1,a-1},\ldots,u_{n,a-1},t^{j-1,a-1})$$

$$- F(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-2},u_{j+1,a-1},\ldots,u_{n,a-1},t^{j,a-1})].$$

This equation requires (n+1) words of storage for the $F(\cdot)$, n words of storage for the $e_{j,a-1}$, n comparisons to obtain the sign $[\cdot]$, and n additional boolean operations to obtain the e_{ja} which can then be stored where the $e_{j,a-1}$ were stored. Equation (4-10) requires storage of n Δu_j and $u_{j,a-1}$ and performance of n additions or subtractions to obtain the u_{ja} which can be stored where the $u_{j,a-1}$ were stored. Finally, $F(u_{1,a-1},\ldots,u_{n,a-1},t^{0,a-1})$ must be shifted to the memory location of $F(u_{1,a-2},\ldots,u_{n,a-2},t^{n,a-1})$ in preparation for the next algorithmic cycle.

These requirements are summarized in Table 1 along with the corresponding requirements for the one-at-a-time and standard gradient techniques. The entries below the main table are the storage requirements and computation times per algorithmic cycle for a 3-stage system and a typical process-control computer with a $2 \cdot 10^{-6}$ sec cycle time, $4 \cdot 10^{-6}$ sec add time, $8 \cdot 10^{-6}$ sec multiply time, and a $16 \cdot 10^{-6}$ sec divide and square root time.[16]

The control-constrained requirements are quite similar to the unconstrained requirements, and are summarized in Table 2.

Table 1. Unconstrained Hardware Requirements

		Multiplexed Gradient Technique	One-at-a-Time Technique	Standard Gradient Technique
Words of Storage		8n + 1	3n + 6	4n + 5
Operations	Boolean	2n + 1	5	1
	Add/Sub	3n	2	3n
	Multiply			3n
	Divide			n
	Sq. Root			1
Words of Storage n = 3		25	15	17
Time (10^{-6} sec) n = 3		50	18	174

Table 2. Control-Constrained Hardware Requirements

		Multiplexed Gradient Technique	One-at-a-Time Technique	Standard Gradient Technique
Words of Storage		10n + 1	5n + 6	7n + 5
Operations	Boolean	5n + 1	10	3n + 1
	Add/Sub	4n	4	3n
	Multiply			3n
	Divide			n
	Sq. Root			1
Words of Storage n = 3		31	21	26
Time (10^{-6} sec) n = 3		80	36	192

The state-constrained requirements, summarized in Table 3, are much greater than the control-constrained requirements, but the computation times are still quite small compared to the time constants (30 seconds to 30 minutes) of the systems being controlled. The parameter k found in Table 3 is the total number of linearized constraints (see 5-15) and the parameter ω is the number of bits in the rounded $d_k^{(r)}$ (see 5-40). The number of operations required by the multiplexed gradient technique is seen to contain a factor 4^{ω}, which means that the computation time is strongly dependent on ω. This dependence is vividly illustrated by the entries below the main table for n=3, k=12, and ω=0, 1, and 2. Of considerable interest, therefore, are the simulation studies of Chapter VIII which establish the ω generally required for convergence to the optimum.

The three techniques compared in Tables 1-3 will be considered again in Chapter VIII. It should be noted here, however, that the apparent superiority of the one-at-a-time technique, as judged by its minimal hardware requirements, is due to a serious shortcoming of the method. Specifically, it cannot guarantee convergence to a constrained optimum even under the assumption of a linear time-invariant system. Consider, for example, a 2-stage system whose performance surface and constraint boundaries are as shown in Figure 19. The one-at-a-time technique is constrained to move parallel to the coordinate axes and, for the initial point shown, can move in none of the four coordinate directions: (1) and (2) result in constraint violations and (3) and (4) result in decreased $F(u_{1a},u_{2a},t^a)$. The one-at-a-time technique, then,

Table 3. State-Constrained Hardware Requirements

		Multiplexed Gradient Technique	One-at-a-Time Technique	Standard Gradient Technique
Words of Storage		$n^2 + 11n + 12 + \frac{3}{2}nk + \frac{7}{2}k + (n+1)2^\omega$	$3n + 6 + 2k$	$n^2 + 5n + 11 + \frac{1}{2}nk + \frac{7}{2}k$
Operations	Boolean	$2n^2 + 3n + 5 + \frac{5}{2}nk + \frac{3}{2}k + n^2 2^\omega + n^2(n-1)4^\omega$	$\frac{3}{2}k + 5$	$1 + \frac{3}{2}nk + k2^n$
	Add/Sub	$3n^3 - 2n^2 + 5n - 2 + 3nk + \frac{3}{2}n^2k + n(n-2)2^\omega + n^4 4^\omega$	$k + 3$	$3n^3 - 2n^2 + n + 1 + nk + k + \frac{3}{2}n^2k + k(n-1)2^n$
	Multiply	$5n^3 - 2n^2 + 2n + 6nk + \frac{3}{2}n^2k + (3n^2 - n + 1)2^\omega + n^3(n-1)4^\omega$		$5n^3 - 2n^2 + 3n - 1 + 4nk + nk2^n + \frac{3}{2}n^2k$
	Divide			n
	Sq. Root			1
Words of Storage $n = 3,\ k = 12$		$\omega = 0$: 154 $\omega = 1$: 158 $\omega = 2$: 166	39	95
Time (10^{-6} sec) $n = 3,\ k = 12$		$\omega = 0$: 6694 $\omega = 1$: 9300 $\omega = 2$: 19264	106	7994

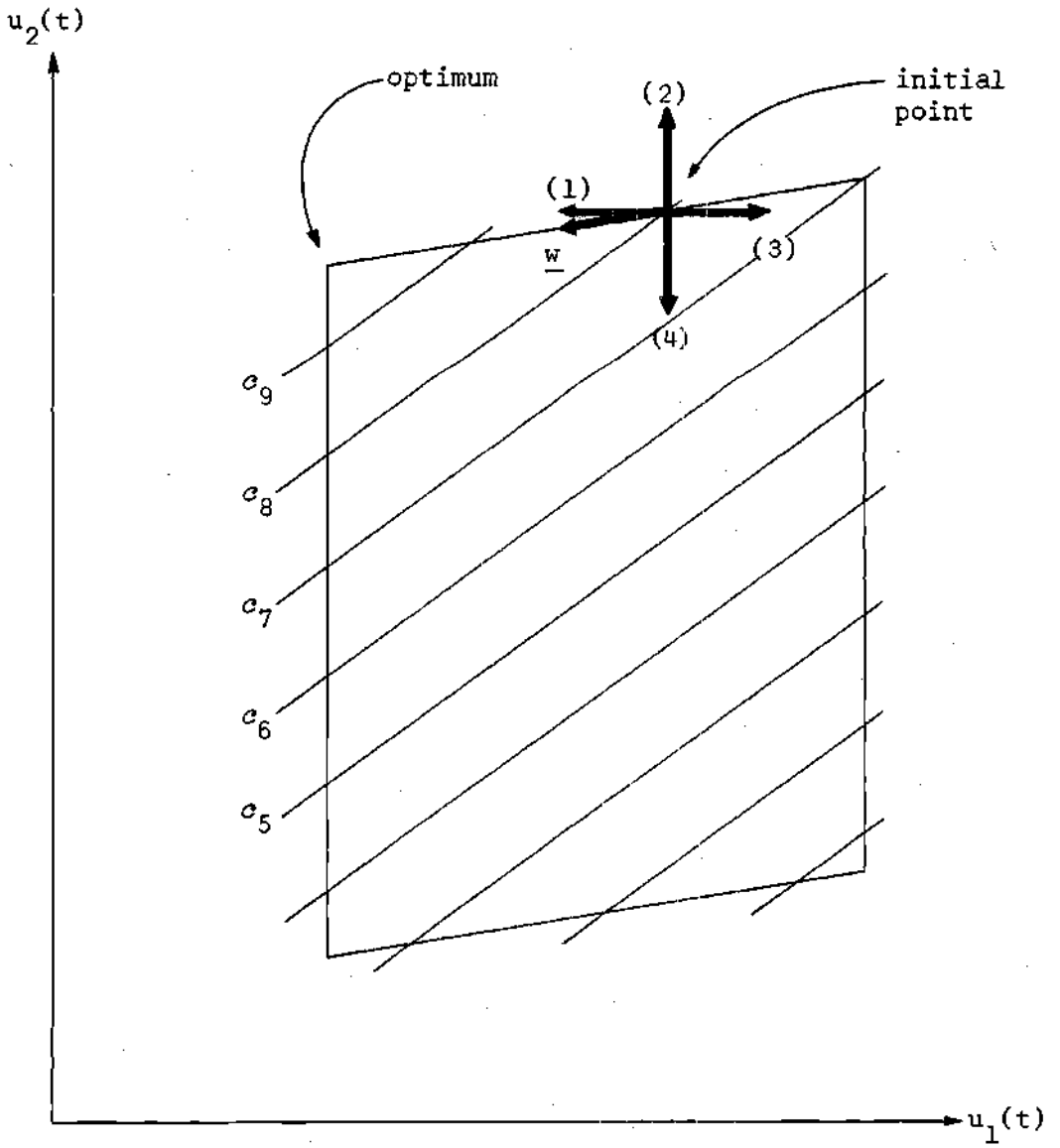


Figure 19. Contour Map of the Function $F(u_{1a},u_{2a},t^a)$

will not converge to the example optimum. In contrast, the multiplexed and standard gradient techniques establish the projected gradient $\underline{w}$ and follow that vector to the constrained optimum.

Word Length Requirements

Consider finally the word length requirements of the multiplexed gradient technique. Measured system data will rarely contain more than 8 bits of information. For unconstrained and control-constrained systems, the arithmetic operations are so few (the frequently encountered case of $\mu=1$ and $F(u_{1a},\ldots,u_{na},t^a) = H_1(u_{1a},\ldots,u_{na},t^a)$ requiring only comparison of measured system data) that an 8-bit machine should be quite adequate.

For state-constrained systems, the critical sequence of computations is that beginning with the measured system data and ending with the $\underline{c}_k$ of (5-22) and the d_k of (5-24). The longest chain of serial computations in this sequence can be shown to be that for the d_k, involving $2(n-1)^2$ additions, n+1 subtractions, and 3n-2 multiplications. Assuming a word length of B_m bits, initial data with a potential error of 1 least-significant-bit, and principal source of cumulative error the 3n-2 multiplications, the number of reliable bits in the d_k can be shown to be

$$B_m - \left[\log_2\left(7 + 12\sum_{j=0}^{n-3} 3^j\right)\right] - 1 \qquad n\geq 2 \tag{7-4}$$

where $[\cdot]$ is the greatest integer function. The derivation of this result is based on the formula

$$(x+\Delta x)(y+\Delta y) = xy\left(1 + \frac{\Delta x}{x} + \frac{\Delta y}{y} + \cancel{\frac{\Delta x \Delta y}{xy}}\right) \quad \text{Neglect} \qquad (7\text{-}5)$$

establishing the per cent error in a product xy as the sum of the per cent errors $\frac{\Delta x}{x}$ and $\frac{\Delta y}{y}$ in the factors x and y. The derivation also assumes that the magnitude of each of the 3n-2 intermediate results is on the order of 2^{B_m}. These assumptions should reasonably approximate floating-point multiplication of data with B_m mantissa bits.

Since $(\omega+1)$ reliable d_k bits are required to establish reliable $d_k^{(r)}$, it follows from (7-4) that the inequality

$$B_m - \left[\log_2\left(7 + 12 \sum_{j=0}^{n-3} 3^j\right)\right] - 1 \geq \omega + 1 \qquad n \geq 2 \qquad (7\text{-}6)$$

must be satisfied. Given in Table 4 is a tabulation of the minimum B_m satisfying (7-6). For n=3 and ω=1, for example, the minimum B_m is 7. It would be prudent, of course, to use a B_m several bits greater than the minimum.

In summary, the storage, computation time, and word length requirements of the multiplexed gradient technique are small enough to permit implementation on any of a number of process-control computers.[16]

Table 4. Minimum B_m for Reliable $d_k^{(r)}$

n = Number of System Stages	ω = Number of Bits in the $d_k^{(r)}$				
	0	1	2	3	4
2	4	5	6	7	8
3	6	7	8	9	10
4	7	8	9	10	11
5	9	10	11	12	13
6	10	11	12	13	14

CHAPTER VIII

ANALYSIS OF OPERATIONAL CHARACTERISTICS

The operational characteristics of the multiplexed gradient technique were investigated by means of digital simulation. Using tubular and stirred-tank example systems, the performance of the technique was compared to that of the one-at-a-time and standard gradient techniques. The criteria of comparison were (a) speed in achieving the optimum, (b) ability to converge to a constrained optimum, (c) ability to track a drifting optimum, and (d) ability to function in a noise-corrupted environment.

Unconstrained Stirred-Tank System

First, a two-stage stirred-tank system used by Schindler and Aris[19] was considered. Principal components of the system, shown in Figure 20, were the two stirred tanks, a measuring instrument, two set point controllers, and the process-control computer. Within the two tanks, the four species A_1, A_2, A_3, and A_4 were undergoing the reversible second-order reaction

$$A_3 + A_4 \rightleftharpoons A_1 + A_2. \qquad (8\text{-}1)$$

It was desired to maintain the tank temperatures at such values as would favor the production of A_1. To this end, the reactant temperatures were

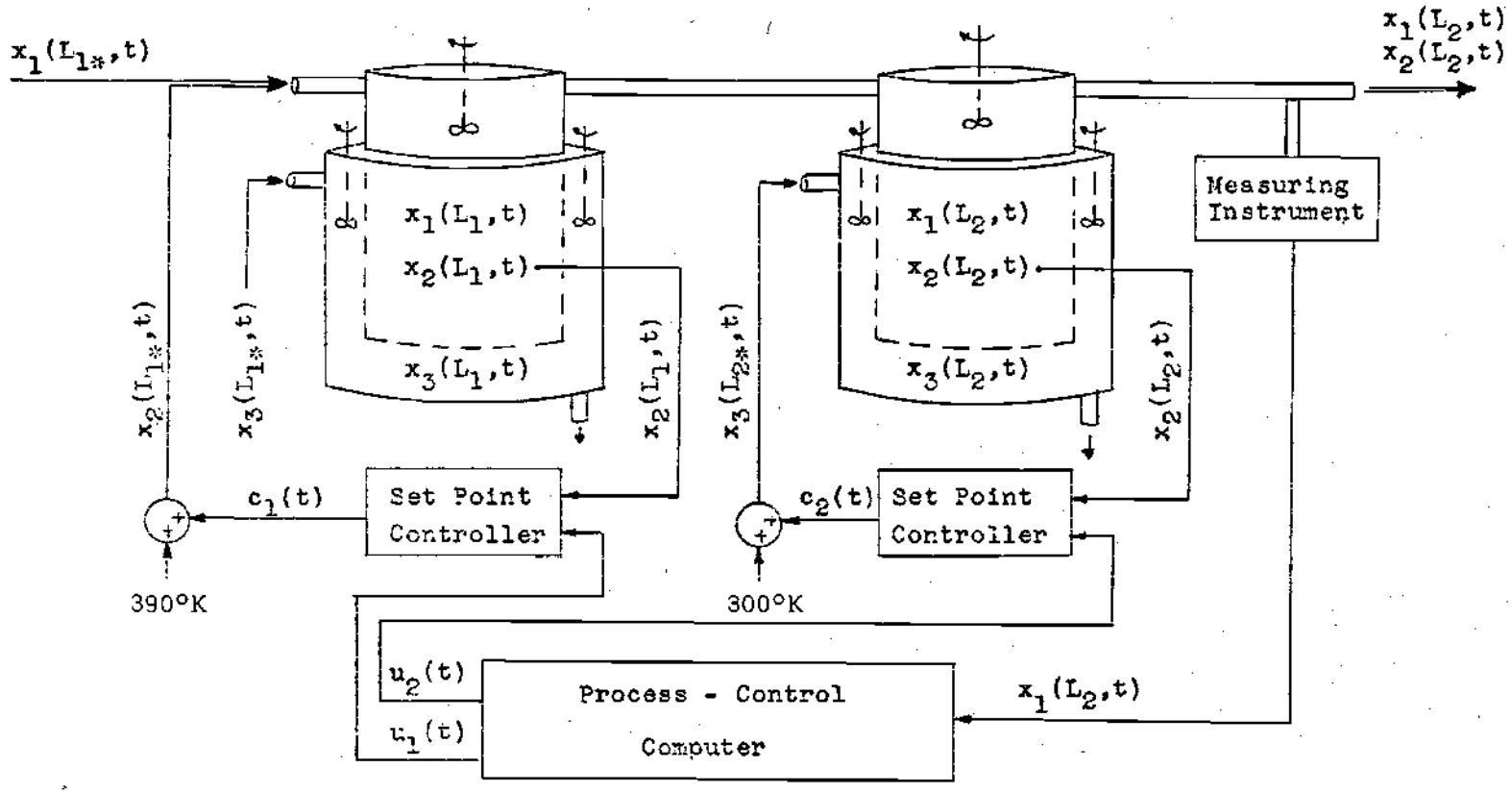


Figure 20. Unconstrained System with Extremum Controller

controlled by proportional-plus-integral set point controllers operating, in the first tank, on the inlet reactant temperature and, in the second tank, on the inlet coolant temperature.[48,49] Periodically, the measuring instrument sampled the exit stream and established the extent of the reaction (8-1). Then, the process-control computer made adjustments in the temperature set points to increase the reaction extent.

The equations describing the example system, not known to the process-control computer but necessary in the simulation of the system, were of the degenerate form given by (2-12). Specifically, with $x_1(\ell,t)$ the reaction extent, $x_2(\ell,t)$ the reactant temperature, and $x_3(\ell,t)$ the coolant temperature, the system was described by the ordinary differential equations[19]

$$\frac{dx_1(L_j,t)}{dt} = \frac{1}{\theta_j}[x_1(L_{j*},t)-x_1(L_j,t)] + r[x_1(L_j,t),x_2(L_j,t)] \quad (8\text{-}2)$$

$$\frac{dx_2(L_j,t)}{dt} = \frac{1}{\theta_j}[x_2(L_{j*},t)-x_2(L_j,t)] - \frac{H_j}{\theta_j}[x_2(L_j,t)-x_3(L_j,t)] \quad (8\text{-}3)$$

$$+ J\times r[x_1(L_j,t),x_2(L_j,t)]$$

$$\frac{dx_3(L_j,t)}{dt} = \frac{1}{\theta_{cj}}[x_3(L_{j*},t)-x_3(L_j,t)] - \frac{K_jH_j}{\theta_{cj}}[x_3(L_j,t)-x_2(L_j,t)] \quad (8\text{-}4)$$

where the kinetic rate expression was[19]

$$r[x_1(L_j,t),x_2(L_j,t)] \tag{8-5}$$

$$= A_f e^{-E_f/\left(R\times x_2(L_j,t)\right)} \left(c_{30}-x_1(L_j,t)\right)\left(c_{40}-x_1(L_j,t)\right)$$

$$- A_r e^{-E_r/\left(R\times x_2(L_j,t)\right)} \left(c_{10}+x_1(L_j,t)\right)\left(c_{20}+x_1(L_j,t)\right).$$

The six stage inlet conditions were[19]

$$x_1(L_{1*},t) = 0.0 \text{ mol/liter} \tag{8-6}$$

$$x_2(L_{1*},t) = \text{stage 1 control}$$

$$x_3(L_{1*},t) = 300^\circ K$$

$$x_1(L_{2*},t) = x_1(L_1,t)$$

$$x_2(L_{2*},t) = x_2(L_1,t)$$

$$x_3(L_{2*},t) = \text{stage 2 control.}$$

The set point controllers, shown schematically in Figure 21, were the output-saturated proportional-plus-integral devices common in reactor control.[48,49]

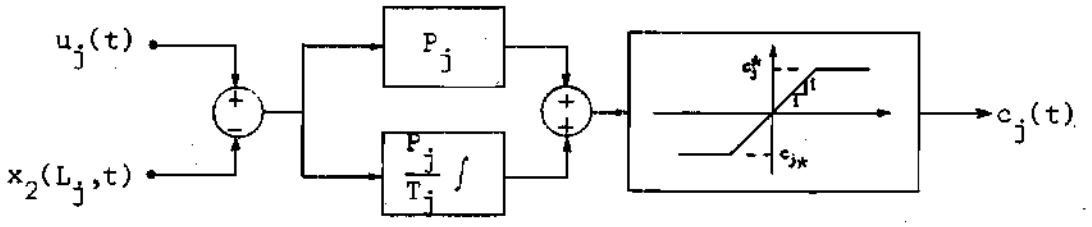


Figure 21. Set Point Controller Configuration

Parameters of the system[19] and set point controllers were as given in Table 5.

Table 5. Stirred-Tank System Parameters

Parameter	Value	Units	Parameter	Value	Units
θ_1	1	min	E_r	21,400	cal/mol
θ_2	2	min	c_{10}	0.0	mol/liter
θ_{c1}	5/12	min	c_{20}	0.0	mol/liter
θ_{c2}	5/12	min	c_{30}	1.0	mol/liter
H_1	0.8	-	c_{40}	1.0	mol/liter
H_2	1.6	-	P_1	2.0	-
K_1	5/12	-	P_2	2.0	-
K_2	5/12	-	T_1	0.1	min
J	6.3	(°K-liter)/mol	T_2	0.5	min
R	1.9869	cal/(°K-mol)	c_{1*}	-90	°K
A_f	1.356×10^{12}	liter/(mol-min)	c_{2*}	-50	°K
A_r	2.346×10^{13}	liter/(mol-min)	c_1^*	60	°K
E_f	18,250	cal/mol	c_2^*	100	°K

In comparing (8-2)-(8-5) to (2-12), it should be noted that the set point controllers, during periods of steady state, effectively uncouple $x_1(L_j,t)$ from $x_2(L_j,t)$ and $x_3(L_j,t)$. Thus, in the interval specified by (3-51), (8-2) can be approximated by

$$\frac{dx_1(L_j,t)}{dt} = \frac{1}{\theta_j}[x_1(L_{j*},t)-x_1(L_j,t)] + r[x_1(L_j,t),u_j(t)]. \quad (8\text{-}7)$$

This equation was not used in the simulation. It was of interest, however, to note that with $\underline{x}(L_j,t) = x_1(L_j,t)$ it was of the form of (2-12).

Appropriate values for the system parameters τ_{js} (see 3-40) and controller parameters τ_{jc} (see 2-14) were found by preliminary simulation to be

$$\tau_{1s} = 1 \text{ min} \qquad \tau_{2s} = 2 \text{ min} \quad (8\text{-}8)$$

$$\tau_{1c} = 1 \text{ min} \qquad \tau_{2c} = 2 \text{ min.}$$

Then from (4-18)-(4-21) and (B-10)-(B-13), the sequencing equations

$$t^{2,a-1} = \text{Max}\{t_{1,a-2}+\tau_{1c}+\tau_{1s}+\tau_{2s}, t^{1,a-2}\} \quad (8\text{-}9)$$

$$t_{2,a-1} = t^{2,a-1}$$

$$t_{1,a-1} = \text{Max}\{t_{2,a-1}, t_{2,a-1}+\tau_{2c}+\tau_{2s}, t^{2,a-1}\}$$

$$t^{1,a-1} = t_{1,a-1}$$

were obtained. After rearrangement, these became

$$t^{2,a-1} = t_{1,a-2} + \tau_{1c} + \tau_{1s} + \tau_{2s} \quad (8\text{-}10)$$

$$t_{2,a-1} = t^{2,a-1}$$

$$t_{1,a-1} = t_{2,a-1} + \tau_{2c} + \tau_{2s}$$

$$t^{1,a-1} = t_{1,a-1}.$$

Finally, inserting the parameters of (8-8),

$$t^{2,a-1} = t_{2,a-1} = t_{1,a-2} + 4 \qquad (8\text{-}11)$$

$$t^{1,a-1} = t_{1,a-1} = t_{2,a-1} + 4.$$

Thus, the $u_1(t)$ and $u_2(t)$ set points were adjusted alternately, with one adjustment every 4 minutes.

A contour map of the steady-state $x_1(L_2,t)$ as a function of the $u_1(t)$ and $u_2(t)$ set points is given in Figure 22. Also shown are the paths taken to the optimum, from an initial point of $u_1(t) = 340°K$ and $u_2(t) = 320°K$, by the multiplexed gradient, one-at-a-time, and standard gradient techniques. Set point increments for the multiplexed gradient and one-at-a-time techniques were $\Delta u_1 = \Delta u_2 = 5°K$. Increments for the standard gradient technique were such that the Euclidean distance moved was 5°K. Simulation of the operations performed by the process-control computer, in the case of the multiplexed gradient technique, was through a sub-routine restricting the Burroughs B5500 to an 8-bit fixed-point internal format.

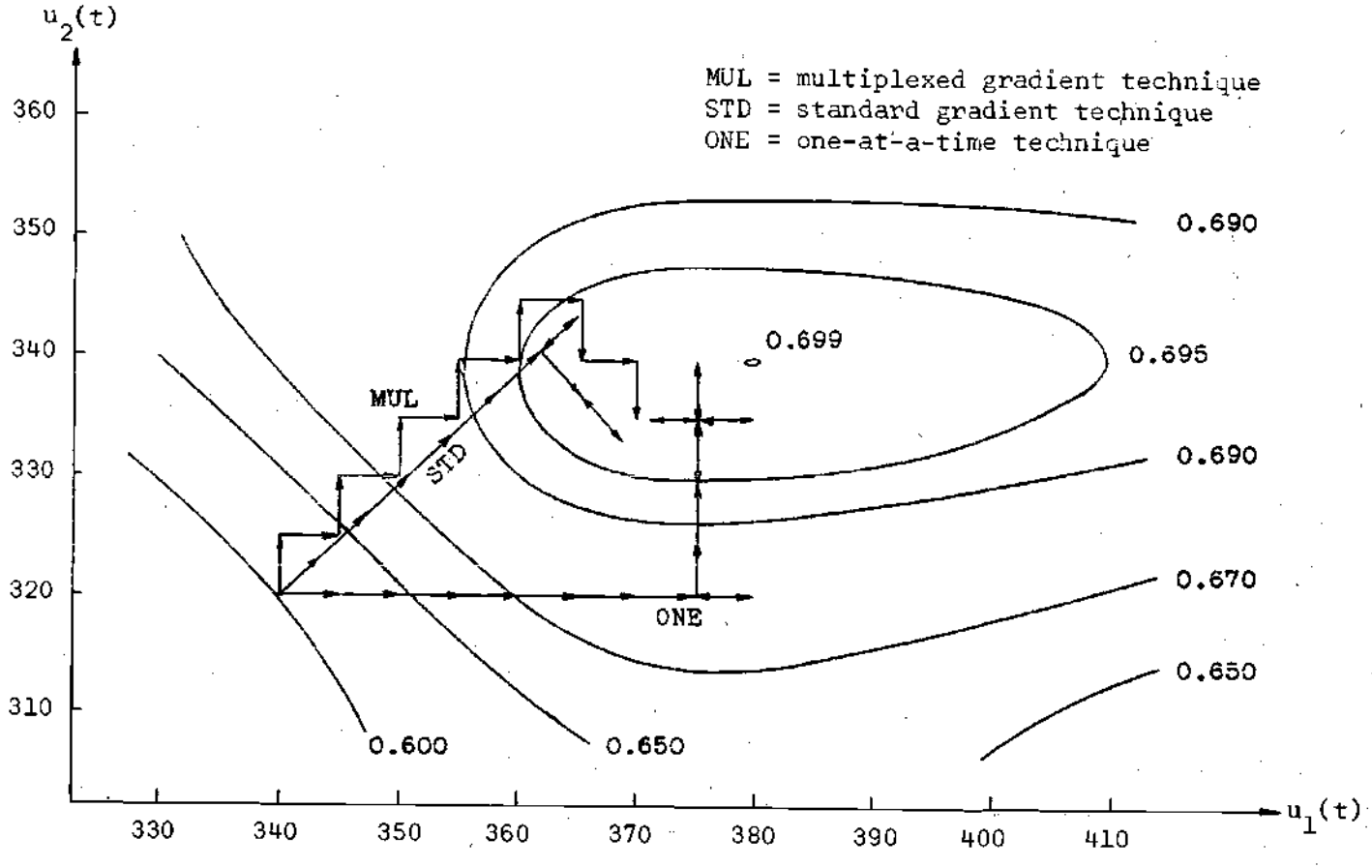


Figure 22. Contour Map of Unconstrained $x_1(L_2,t)$

Considering the pair of moves making one algorithmic cycle, the multiplexed gradient technique may be seen to move in a direction fixed by the ratio $\Delta u_2/\Delta u_1$. In contrast, the one-at-a-time technique moves parallel to the coordinate axes and the standard gradient technique moves in the direction of steepest ascent. The comparative speeds of the three techniques may be seen in Figure 23, where the paths of Figure 22 are plotted as functions of time. Even in this degenerate stirred-tank system, the multiplexed gradient technique was faster in achieving the optimum than the other two techniques. This speed was due to the ability of the multiplexed technique to simultaneously increase $x_1(L_2,t)$ and update the $x_1(L_2,t)$ model.

Control-Constrained Tubular System

Considered next was an example illustrating the control-constrained algorithm. The system simulated, shown in Figure 24, was a two-stage tubular reactor used by Tsai[4] and others[50,51] in various optimization studies. The temperature of the coolant in the jackets surrounding the reactor was controlled by set point controllers whose dynamics were sufficiently fast, compared to those of the reactor, that they could be assumed to be ideal.[4,17,18] The temperature set points, constrained to lie between specified upper and lower bounds, were adjusted by the extremum control algorithm to maximize the exit concentration of the component B in the reaction

$$A \longrightarrow B \longrightarrow C, \qquad (8\text{-}12)$$

where A was the feed, B the desired product, and C a waste product.

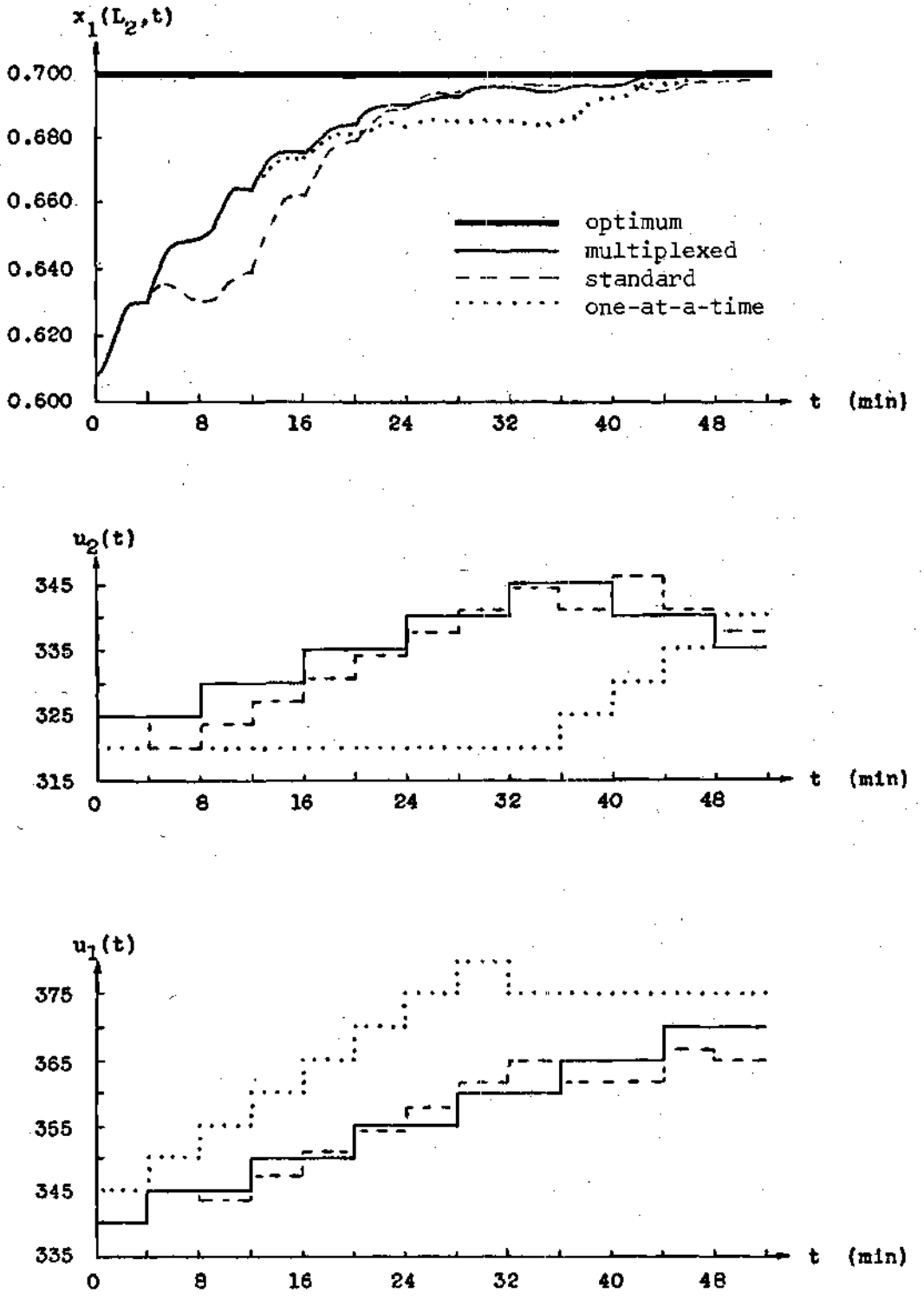


Figure 23. Unconstrained Optimization Paths

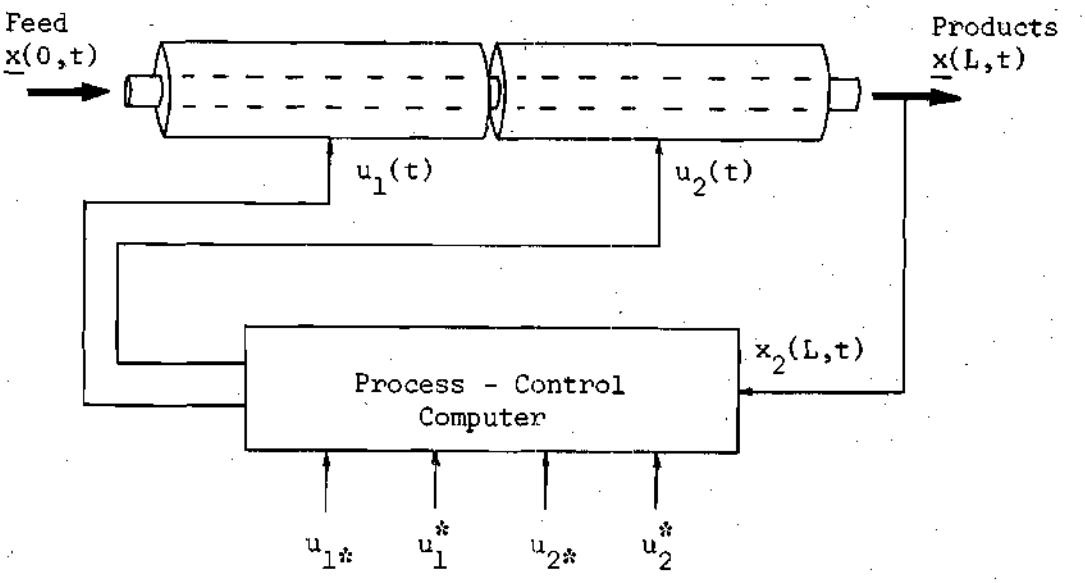


Figure 24. Control-Constrained System with Extremum Controller

The equations describing the tubular reactor were of the basic form given by (2-1). Specifically, with $x_1(\ell,t)$ the concentration of the species A, $x_2(\ell,t)$ the concentration of the species B, and $x_3(\ell,t)$ the reactant temperature, the reactor was described by the partial differential equations[4]

$$\frac{\partial x_1(\ell,t)}{\partial t} + v\,\frac{\partial x_1(\ell,t)}{\partial \ell} = -k_1 x_1(\ell,t) \tag{8-13}$$

$$\frac{\partial x_2(\ell,t)}{\partial t} + v\,\frac{\partial x_2(\ell,t)}{\partial \ell} = k_1 x_1(\ell,t) - k_2 x_2(\ell,t) \tag{8-14}$$

$$\frac{\partial x_3(\ell,t)}{\partial t} + v\,\frac{\partial x_3(\ell,t)}{\partial \ell} = c_3[u(\ell,t)-x_3(\ell,t)] + c_4[-\Delta H_1 k_1 x_1(\ell,t)-\Delta H_2 k_2 x_2(\ell,t)] \tag{8-15}$$

where the kinetic rate constants were[4]

$$k_1 = a_{11} e^{-a_{12}/x_3(\ell,t)}$$
$$k_2 = a_{21} e^{-a_{22}/x_3(\ell,t)} . \tag{8-16}$$

The coolant temperature was given by[4]

$$u(\ell,t) = \begin{cases} u_1(t) & 0<\ell\leq L/2 \\ u_2(t) & L/2<\ell\leq L \end{cases} \tag{8-17}$$

and the system boundary conditions were given by[4]

$$x_1(0,t) = x_{1B} \tag{8-18}$$
$$x_2(0,t) = x_{2B}$$
$$x_3(0,t) = x_{3B}.$$

The controlled coolant temperatures were constrained to satisfy[4]

$$u_{1*} \leq u_1(t) \leq u_1^* \tag{8-19}$$
$$u_{2*} \leq u_2(t) \leq u_2^*.$$

System parameters[4] were as given in Table 6.

Table 6. Control-Constrained Tubular Reactor Parameters*

Parameter	Value	Units	Parameter	Value	Units
L/v	0.6	hr	a_{22}	4,400	°K
c_3	20	1/hr	x_{1B}	0.95	-
c_4	0.001	(°K-gm mole)/cal	x_{2B}	0.05	-
ΔH_1	-50,000	cal/gm mole	x_{3B}	350	°K
ΔH_2	-200,000	cal/gm mole	u_{1*}	320	°K
a_{11}	1,280,000	1/hr	u_{2*}	320	°K
a_{12}	4,900	°K	u_1^*	380	°K
a_{21}	320,000	1/hr	u_2^*	380	°K

*This example is often found in the literature without units. For completeness, the above set of units[35] were appended to the parameters given in the reference.[4]

The sequencing equations

$$t^{2,a-1} = \text{Max}\{t_{1,a-2} + \frac{L}{2v} + \frac{L}{2v},\ t^{1,a-2}\} \tag{8-20}$$

$$t_{2,a-1} = t^{2,a-1}$$

$$t_{1,a-1} = \text{Max}\{t_{2,a-1},\ t_{2,a-1} + \frac{L}{2v} - \frac{L}{2v},\ t^{2,a-1} - \frac{L}{2v}\}$$

$$t^{1,a-1} = t_{1,a-1} + \frac{L}{2v}$$

were obtained from (4-18)-(4-21) and (B-10)-(B-13). After rearrangement, these became

$$t^{2,a-1} = t_{1,a-2} + \frac{L}{v} \qquad (8\text{-}21)$$

$$t_{2,a-1} = t^{2,a-1}$$

$$t_{1,a-1} = t_{2,a-1}$$

$$t^{1,a-1} = t_{1,a-1} + \frac{L}{2v}.$$

Finally, inserting L/v from Table 6,

$$t_{1,a-1} = t_{2,a-1} = t_{1,a-2} + 0.6 \qquad (8\text{-}22)$$

$$t^{2,a-1} = t_{1,a-1}$$

$$t^{1,a-1} = t_{1,a-1} + 0.3.$$

Thus, the $u_1(t)$ and $u_2(t)$ set points were adjusted simultaneously every 0.6 hours. The $x_2(L,t)$ measurements were made alternately every 0.3 hours. This sequencing should be compared to that given in (8-11) for the degenerate system, where the set points were adjusted alternately.

A contour map of the steady-state $x_2(L,t)$ as a function of the $u_1(t)$ and $u_2(t)$ set points is given in Figure 25. The bounds on the map are the upper and lower control constraints of (8-19). Shown on the map are the paths taken to the optimum, from an initial point of $u_1(t) = 330°K$ and $u_2(t) = 360°K$, by the multiplexed gradient, one-at-a-time (first few moves only), and standard gradient (first few moves only) techniques. Set point increments for the multiplexed gradient and one-at-a-time techniques were $\Delta u_1 = \Delta u_2 = 2.5°K$. Increments for the standard gradient technique were $\sqrt{2.5^2+2.5^2} = 3.53°K$ so that the speeds of the standard and multiplexed gradient techniques could be equitably compared.

Simulation of the operations performed by the process-control computer, in the case of the multiplexed gradient technique, was through the subroutine restricting the Burroughs B5500 to an 8-bit fixed-point internal format.

The partial differential equations describing the tubular reactor were simulated by transforming them into ordinary differential equations. Using the procedure of (3-1)-(3-6), the t_α of (3-2) were the discrete times at which fluid elements were assumed to enter the reactor. In this example, these times were separated by 0.15 hours, which resulted in four fluid elements being in the reactor at all times.

The comparative speeds and convergence properties of the three techniques may be seen in Figure 26, where the paths of Figure 25 are plotted as functions of time. The standard gradient technique failed to converge to the optimum, a result not unexpected in view of previous

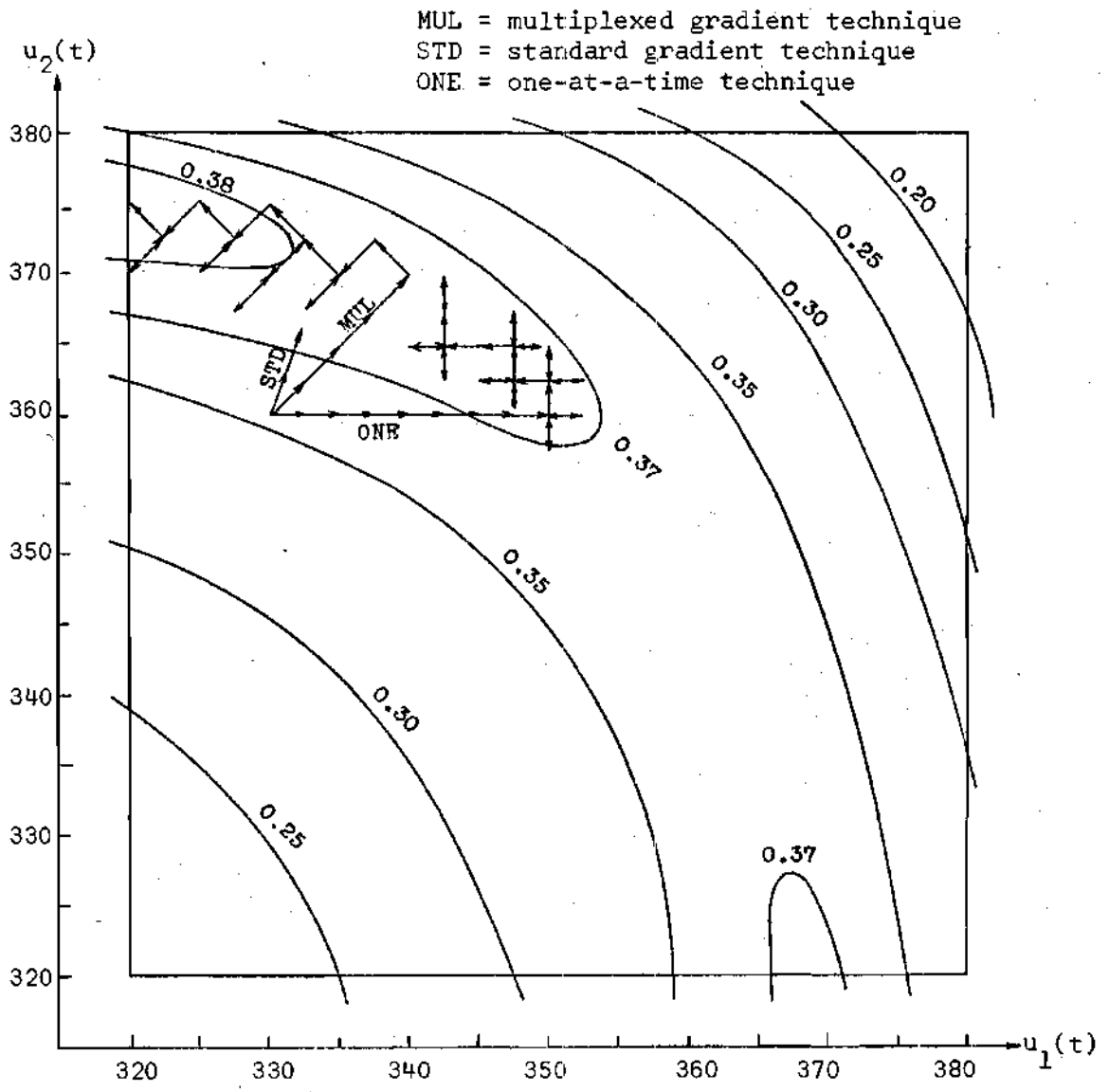


Figure 25. Contour Map of Control-Constrained $x_2(L,t)$

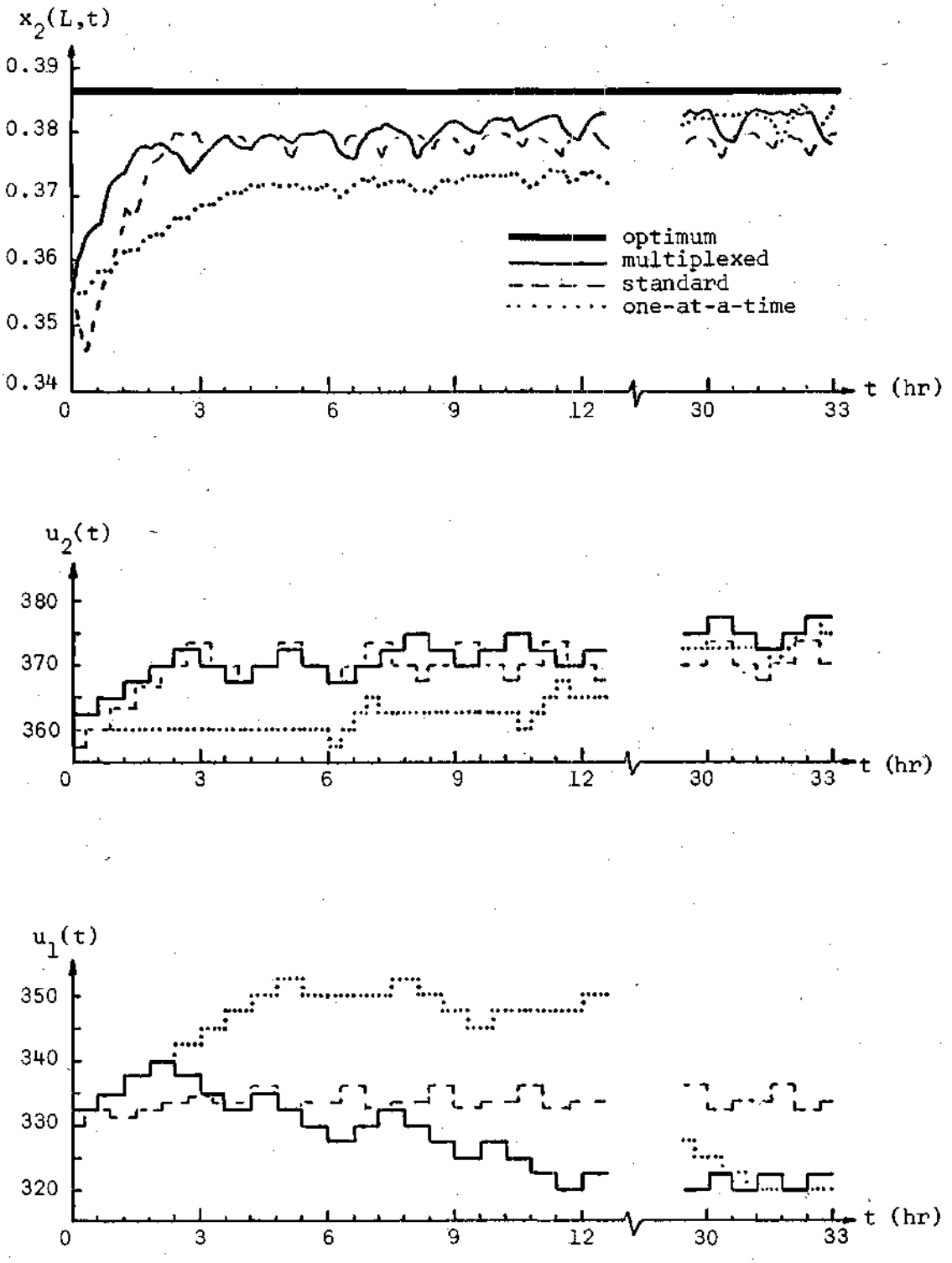


Figure 26. Control-Constrained Optimization Paths

researcher's experience with this technique on curved ridges.[47] The one-at-a-time technique converged to the optimum, but only after 31.5 hours. The multiplexed gradient technique converged in less than half this time, namely 12.0 hours. In this example, the superior technique was clearly the multiplexed gradient.

Two-Stage State-Constrained System

The example chosen to illustrate the state-constrained algorithm was the tubular reactor considered above, but with control constraints deleted and state constraints added. Specifically, the reactant temperature at points $\ell = \frac{1}{4} L$, $\frac{1}{2} L$, $\frac{3}{4} L$, and L was constrained lie between specified upper and lower bounds. In the notation of (2-18), these constraints were

$$z_{jkr*} \leq x_3(\ell_{jk},t) \leq z^*_{jkr} \qquad (8\text{-}23)$$

$$j=1,2 \qquad k=1,2 \qquad r=1$$

where

$$\ell_{jk} = \frac{L}{4} + (j-1)\frac{L}{2} + (k-1)\frac{L}{4}. \qquad (8\text{-}24)$$

The state-constrained system is shown in Figure 27. System equations were those previously given as (8-13)-(8-18), and system parameters, differing slightly from those of Table 6, were as given in Table 7.

As noted above (5-12), state constraints cause no change in the set point command times $t_{j,a-1}$ or measurement sample times $t^{j,a-1}$.

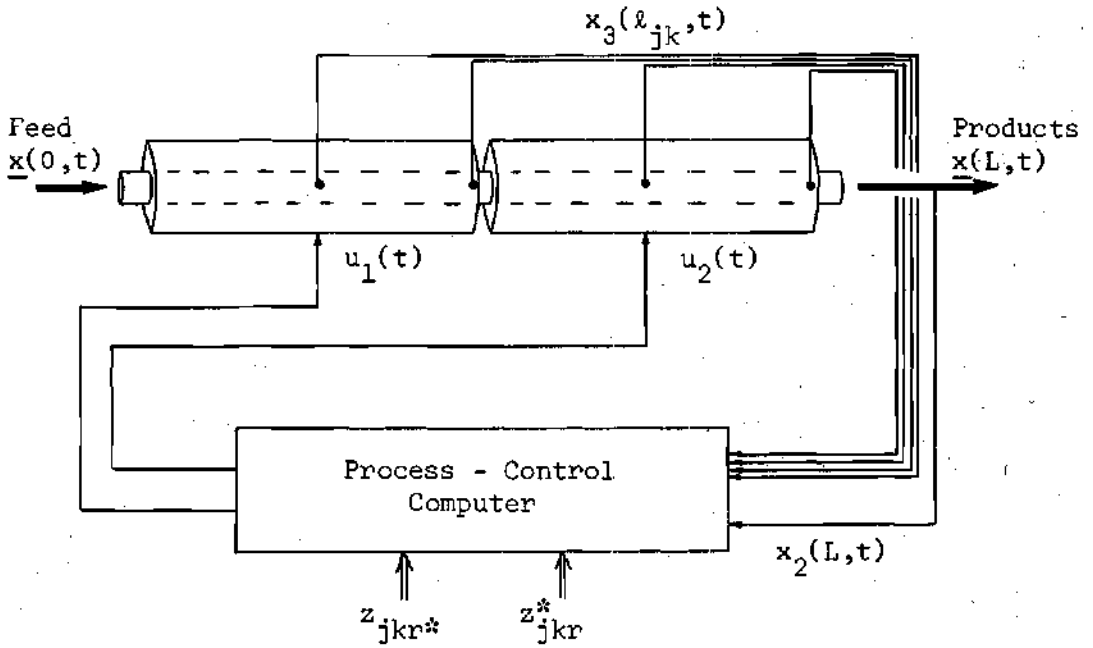


Figure 27. Two-Stage State-Constrained System

Table 7. Two-Stage State-Constrained Tubular Reactor Parameters

Parameter	Value	Units	Parameter	Value	Units
L/v	0.6	hr	x_{2B}	0.05	-
c_3	20	1/hr	x_{3B}	350	°K
c_4	0.001	(°K-gm mole)/cal	z_{111*}	300	°K
ΔH_1	-50,000	cal/gm mole	z_{121*}	300	°K
ΔH_2	-200,000	cal/gm mole	z_{211*}	300	°K
a_{11}	640,000	1/hr	z_{221*}	300	°K
a_{12}	4,900	°K	z^*_{111}	400	°K
a_{21}	320,000	1/hr	z^*_{121}	400	°K
a_{22}	7,200	°K	z^*_{211}	400	°K
x_{1B}	0.95	-	z^*_{221}	400	°K

These, therefore, were as given by (8-22). The $x_3(\ell_{jk},t)$ measurement times were as given by (5-12). Specifically, with j=1,

$$t^{(1k)1,a-1} = t_{2,a-1}, \tag{8-25}$$

and with j=2, after inserting (B-10)-(B-13),

$$t^{(2k)2,a-1} = t_{2,a-1} \tag{8-26}$$

$$t^{(2k)1,a-1} = t_{1,a-1} + 0.15k.$$

Equation (8-25) established the times at which measurements were made for the $x_3(\ 1/4\ L,t)$ and $x_3(1/2\ L,t)$ locally-linear models, and (8-26) established the times at which the measurements were made for the corresponding $x_3(3/4\ L,t)$ and $x_3(L,t)$ models.

A contour map of the steady-state $x_2(L,t)$ as a function of the $u_1(t)$ and $u_2(t)$ set points is given in Figure 28. The contours of constant $x_2(L,t)$ differ from those of Figure 25 because the two examples had different values for the parameters a_{11} and a_{22}. The bounds on the map result from the state constraints of (8-23). The left and right boundaries are parallel to the $u_2(t)$ axis because they result from constraints on $x_3(1/4\ L,t)$ and $x_3(1/2\ L,t)$, which are independent of $u_2(t)$. The top and bottom boundaries are slightly curved because they result from constraints on $x_3(3/4\ L,t)$ and $x_3(L,t)$, which depend on both $u_1(t)$ and $u_2(t)$. The origin of the constraint boundaries may be more

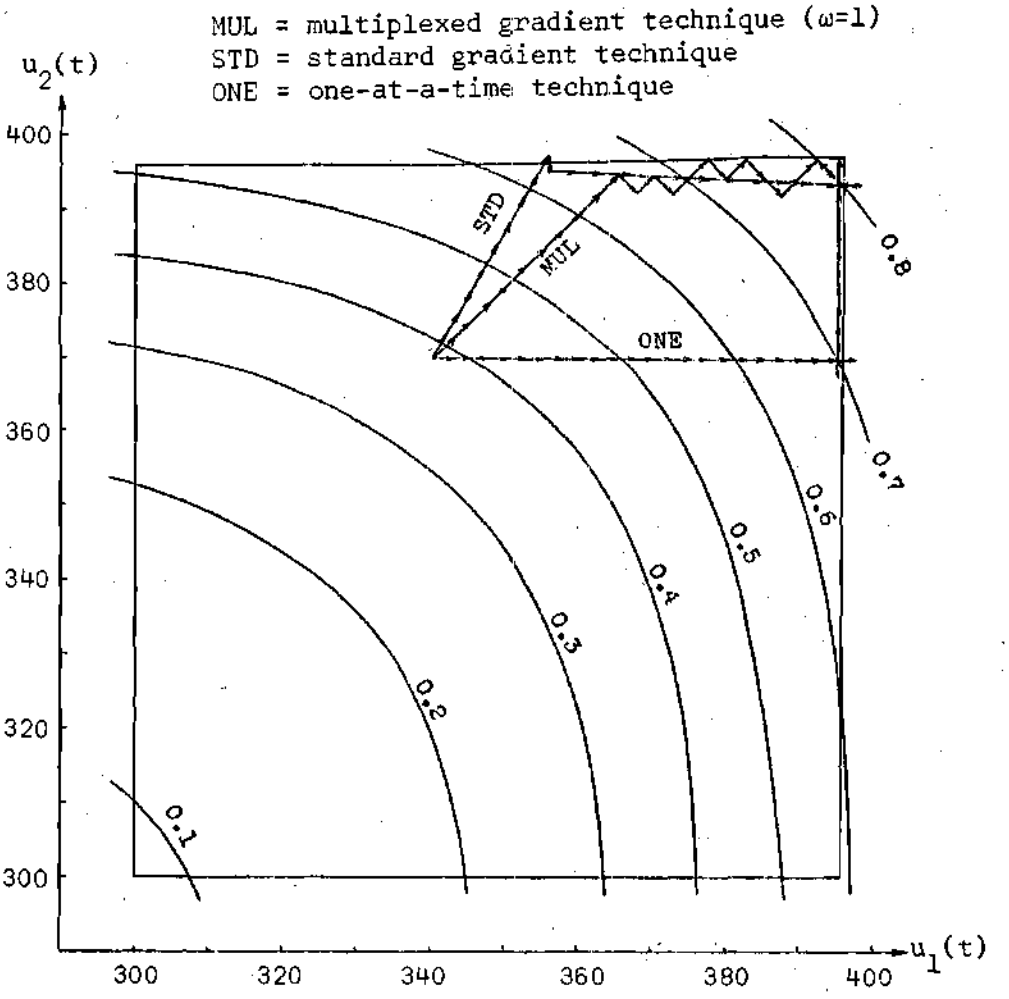


Figure 28. Contour Map of Two-Stage State-Constrained $x_2(L,t)$

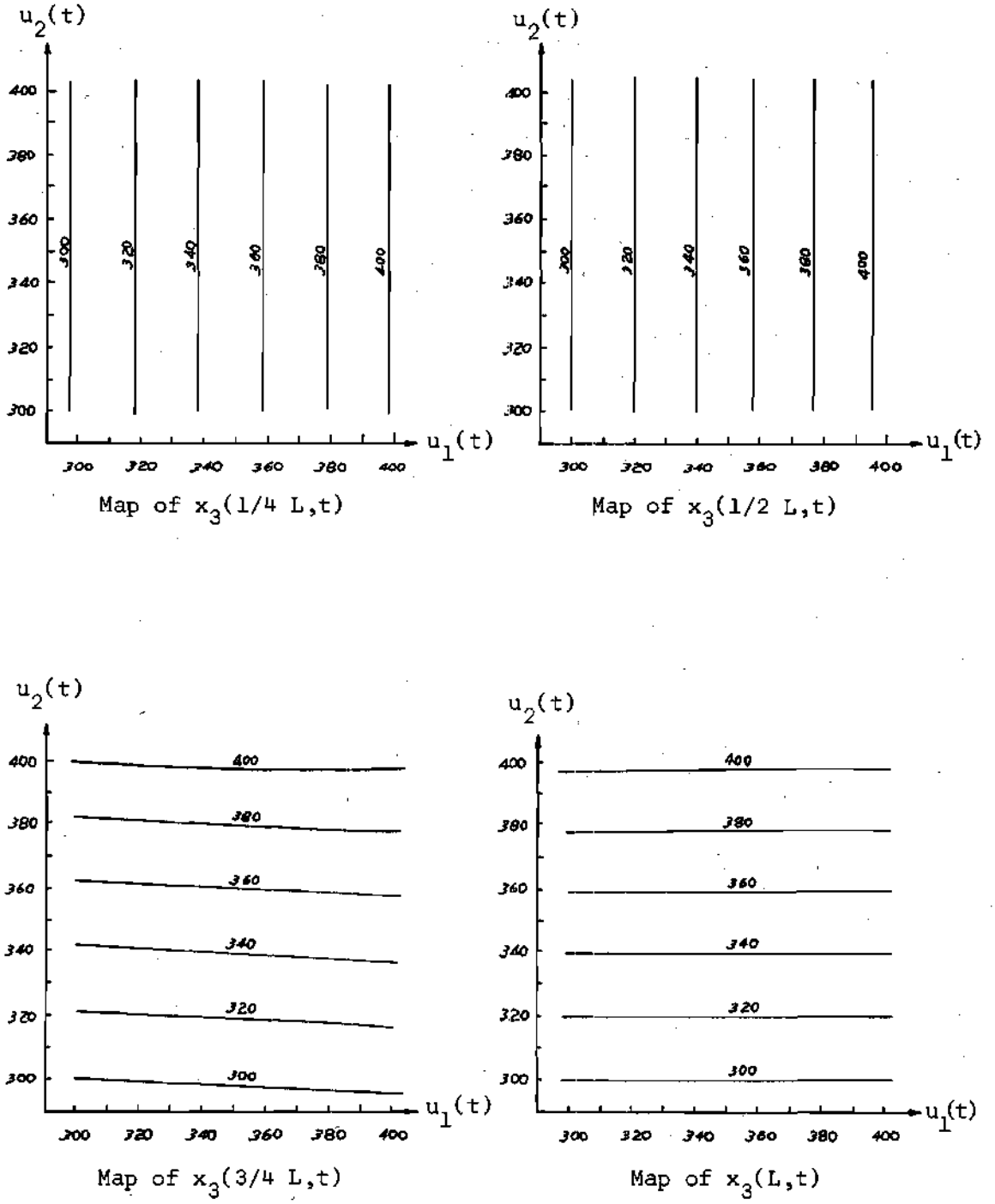


Figure 29. Contour Maps of the Constrained $x_3(\ell_{jk},t)$

clearly seen in the contour maps of steady-state $x_3(\ell_{jk},t)$ given in Figure 29. The linearity of these maps is typical of reactors with good heat transfer between reactant and coolant.

Shown on the map of Figure 28 are the paths taken to the optimum, from an initial point of $u_1(t)$ = 340°K and $u_2(t)$ = 370°K, by the multiplexed gradient, one-at-a-time, and standard gradient (in conjunction with the projected gradient) techniques. Simulation of the operations performed by the process-control computer, in the case of the multiplexed gradient technique, was through a sub-routine restricting the Burroughs B5500 to a 16-bit floating-point format with a B_m of 10. This B_m was several bits greater than the minimum given in Table 4 (page 93), and was found to be quite satisfactory.

The comparative speeds of the three techniques, all of which climbed directly to the optimum, may be seen in Figure 30, where the $x_2(L,t)$ of Figure 28 are plotted as functions of time.

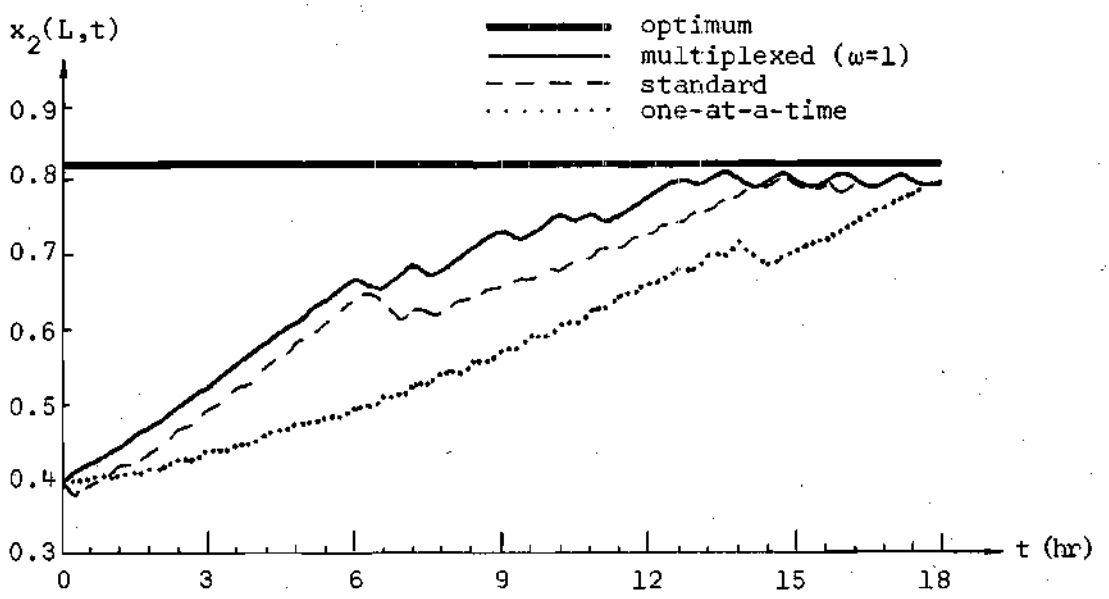


Figure 30. Two-Stage Optimization Response

The multiplexed gradient technique reached the optimum before the other two techniques. In addition, the area under the multiplexed curve was greater, a point of real significance since the dollar value of the species produced was proportional to the integral $\int_0^t x_2(L,t)dt$.

The transient histories of the constrained $x_3(\ell_{jk},t)$ were examined, for the procedure developed in Chapter V, depending in particular on the assumption (5-5), guaranteed constraint satisfaction only during the steady-state portion of the algorithmic cycle. To this end, the partial differential equations describing the tubular reactor were solved by the procedure described earlier, but with 16, rather than 4, fluid elements maintained in the reactor at all times. The results of that simulation are given in Figure 31.

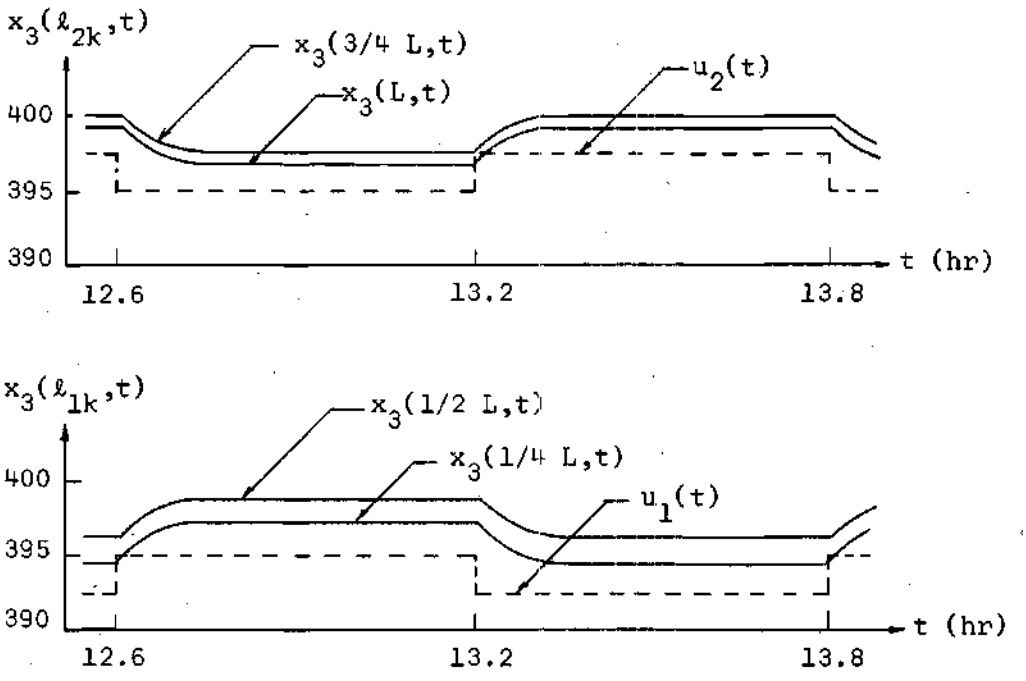


Figure 31. Multiplexed Gradient Transient Histories

The $x_3(\ell_{jk},t)$ shown are for the two algorithmic cycles at the inception of the limit cycle about the optimum, and are typical of those encountered all along the optimization path. The assumption (5-5) is seen to be well justified, for the transient $x_3(\ell_{jk},t)$ remained at or below the 400°K level.

The data presented in Figure 28 was obtained with an ω of 1 (see 5-40). To determine the effect of the parameter ω, simulation runs were made with ω = 0, 1, 2, and 3. An ω of 0 was found to be too small: the optimum was achieved but the path did not follow the upper constraint boundary as it should. Values of 1, 2, and 3 resulted in identical optimization paths, but with increased computation time in the case of ω = 2 and 3. In this example, then, the appropriate value for ω was 1.

Three-Stage State-Constrained System

To further examine the properties of the multiplexed gradient technique, a three-stage example was considered. Its equations, except for the coolant temperature given by

$$u(\ell,t) \begin{cases} u_1(t) & 0<\ell\leq L/3 \\ u_2(t) & L/3<\ell\leq 2L/3 \\ u_3(t) & 2L/3<\ell\leq L, \end{cases} \qquad (8\text{-}27)$$

were as previously given by (8-13)-(8-18). Monitored reactant temperatures, analogous to (8-23), were constrained to satisfy

$$z_{jkr*} \leq x_3(\ell_{jk},t) \leq z^*_{jkr} \qquad (8\text{-}28)$$

$$j=1,2,3 \quad k=1,2 \quad r=1$$

where

$$\ell_{jk} = \frac{L}{6} + (j-1)\frac{L}{3} + (k-1)\frac{L}{6} . \qquad (8\text{-}29)$$

The system was as shown in Figure 32, with parameter values as given in Table 8.

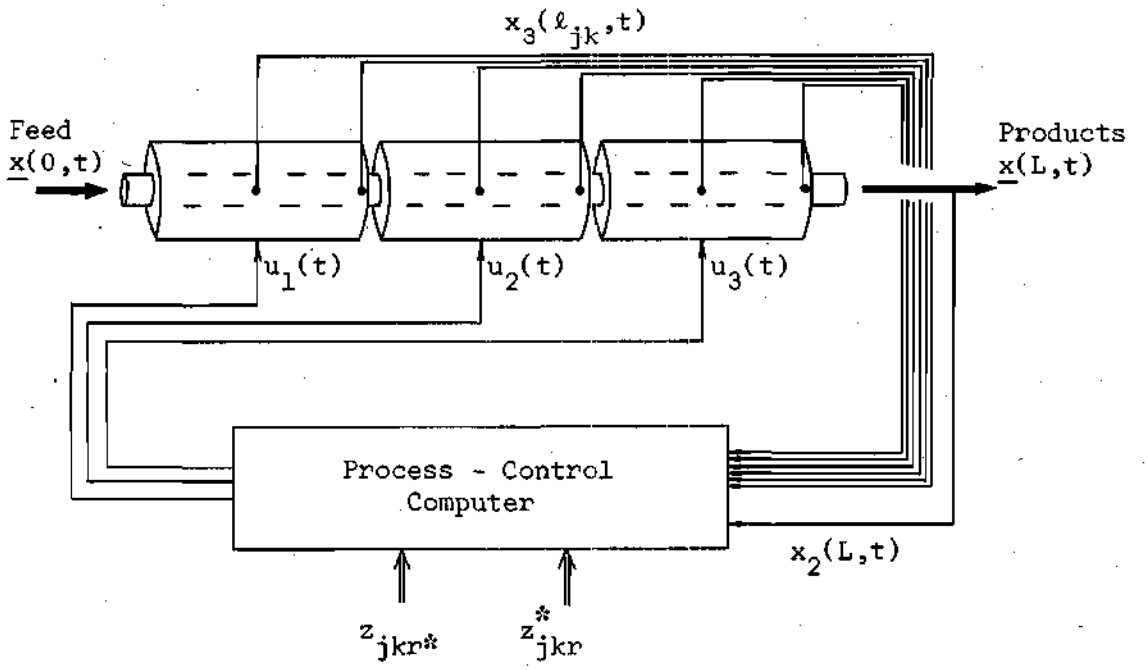


Figure 32. Three-Stage State-Constrained System

In addition to having three stages, this example was of interest because the contours of constant $x_2(L,t)$ were oriented so that the optimum was rather ill-defined. This property of the system may be seen in the sections through the constant $x_2(L,t)$ surfaces given in Figure 33.

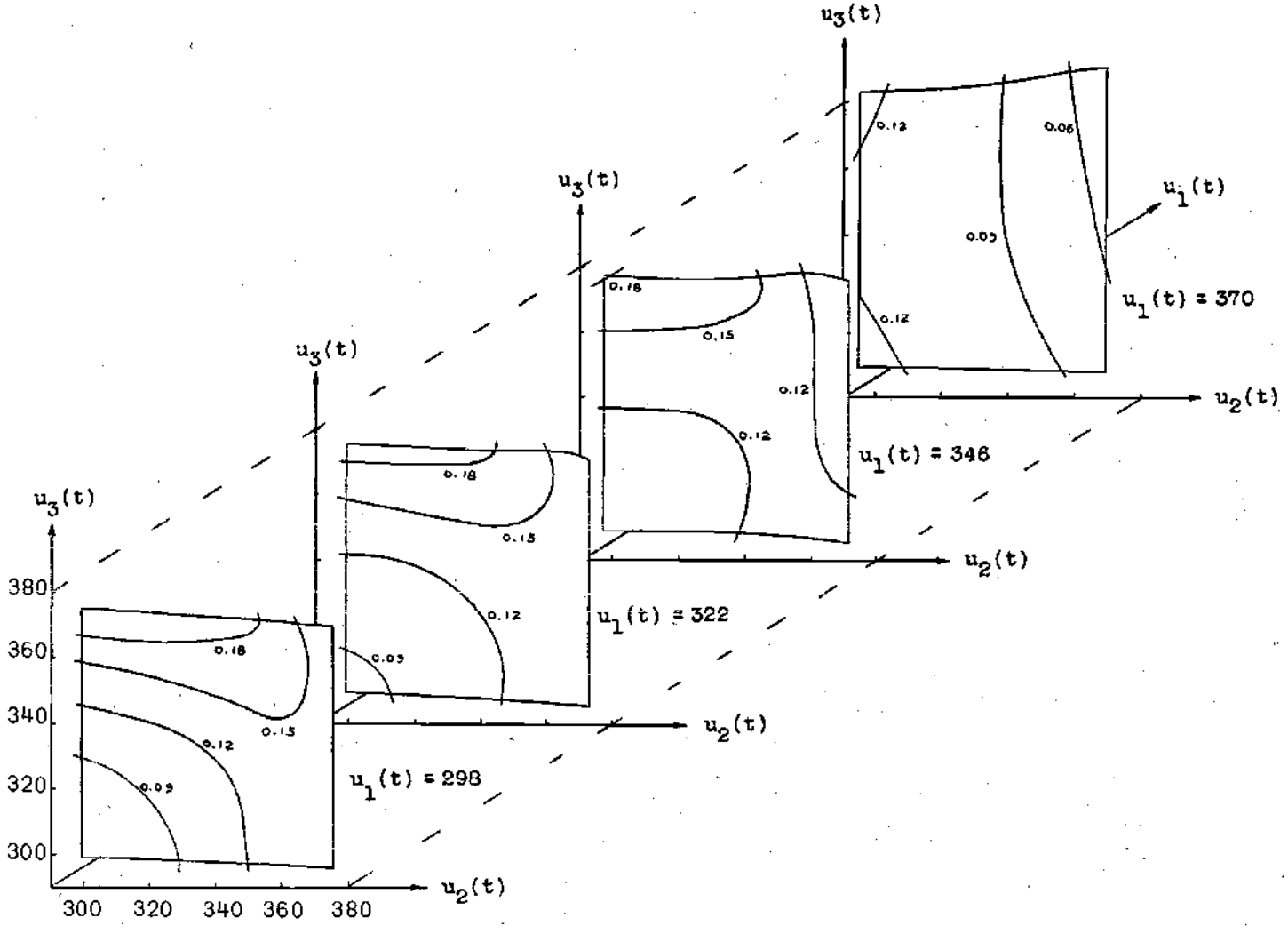


Figure 33. Sections Through $x_2(L,t)$ Surfaces

Table 8. Three-Stage State-Constrained Tubular Reactor Parameters

Parameter	Value	Units	Parameter	Value	Units
L/v	0.6	hr	z_{111}^*	300	°K
c_3	20	1/hr	z_{121}^*	300	°K
c_4	0.001	(°K-gm mole)/cal	z_{211}^*	300	°K
ΔH_1	-50,000	cal/gm mole	z_{221}^*	300	°K
ΔH_2	-200,000	cal/gm mole	z_{311}^*	300	°K
a_{11}	1,280,000	1/hr	z_{321}^*	300	°K
a_{12}	4,900	°K	z_{111}^*	400	°K
a_{21}	14,400	1/hr	z_{121}^*	400	°K
a_{22}	2,800	°K	z_{211}^*	400	°K
x_{1B}	0.95	-	z_{221}^*	400	°K
x_{2B}	0.05	-	z_{311}^*	400	°K
x_{3B}	350	°K	z_{321}^*	400	°K

As indicated in the figure, the constraints of (8-28) were satisfied for $298°K \leq u_1(t) \leq 370°K$ and for $u_2(t)$ and $u_3(t)$ lying within the boundaries shown.

From an initial point of $u_1(t) = 320°K$, $u_2(t) = 330°K$, and $u_3(t) = 365°K$, the multiplexed gradient, one-at-a-time, and standard gradient techniques were applied to the three-stage system. The results are given in Figure 34. As with the two-stage system, the multiplexed and standard gradient techniques were much faster than the one-at-a-time

technique. All three of the techniques did eventually converge to a limit cycle about the constrained optimum.

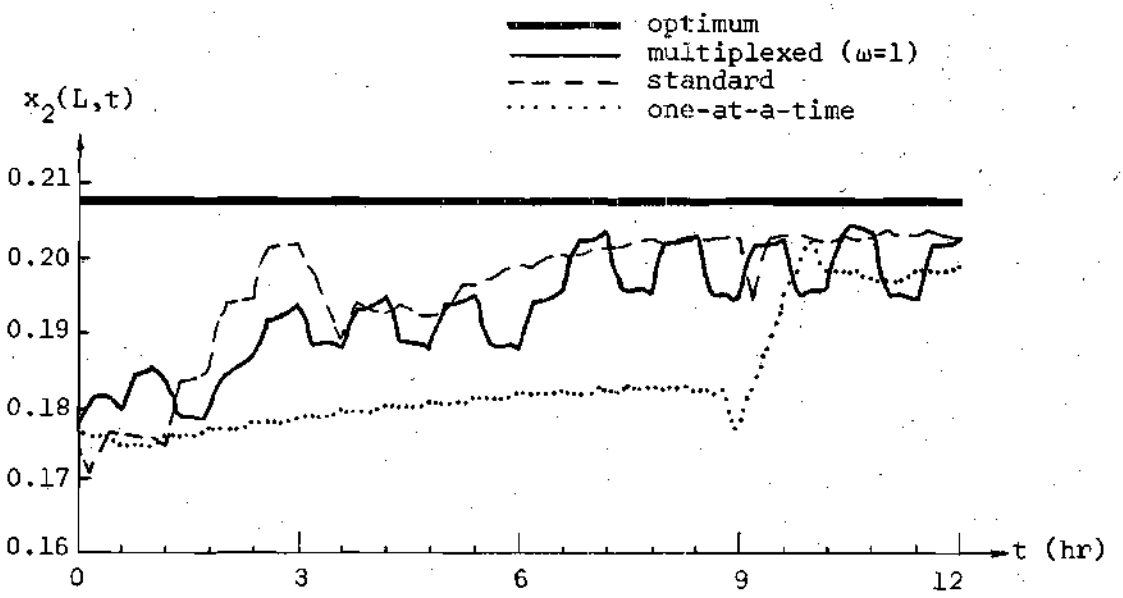


Figure 34. Three-Stage Optimization Response

The effect of the parameter ω was the same as in the two-stage system: ω = 0 was too small, ω = 1 produced good results, and ω = 2 produced results identical to those of ω = 1 except for increased computation time. On the basis of the examples considered, then, an ω of 1 can be recommended.

Time-Varying State-Constrained System

To compare the multiplexed gradient, one-at-a-time, and standard gradient techniques with respect to their ability to track a drifting optimum, the parameters of the reactor of Figure 27 (page 112) were allowed to vary in a manner representative of a changing feedstock.

Specifically, the a_{ij} of Table 7 (page 112) were replaced by

$$a_{11} = 1{,}280{,}000\alpha + 640{,}000\ (1-\alpha) \tag{8-30}$$

$$a_{12} = 4{,}900$$

$$a_{21} = 640{,}000\ \alpha e^{1800(1-\alpha)/350} + 320{,}000\ (1-\alpha)e^{-1800\ \alpha/350}$$

$$a_{22} = 5{,}400\ \alpha + 7{,}200\ (1-\alpha)$$

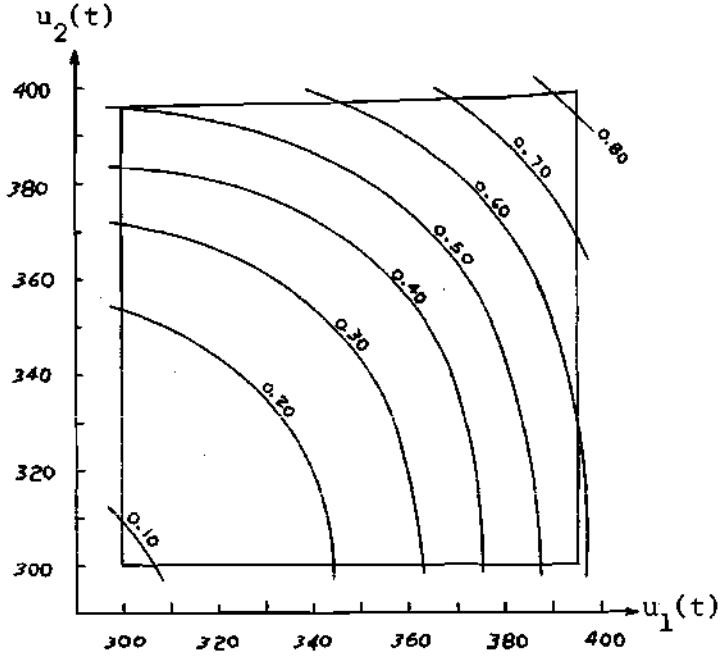
where α was given by

$$\alpha = t/15 \qquad 0 \le t \le 15 \text{ hr.} \tag{8-31}$$

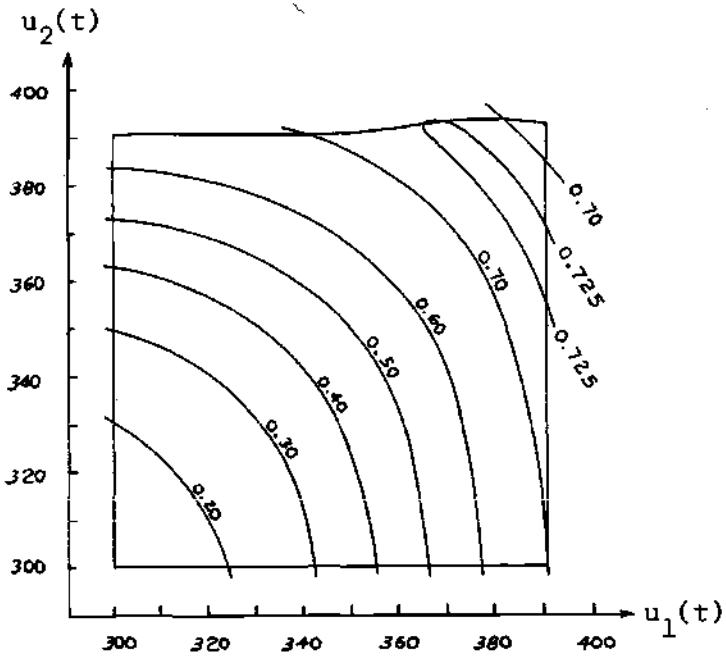
These a_{ij} were used because they produced kinetic rate constants (see 8-16) which were linear in α for $x_3(\ell,t) = 350°K$. This linearity assured a smooth and steady drift in system characteristics over the time interval [0,15]. The initial and final a_{ij} are given in Table 9.

Table 9. Initial and Final Drift Parameters

Parameter	Value at t=0	Value at t=15
a_{11}	640,000	1,280,000
a_{12}	4,900	4,900
a_{21}	320,000	640,000
a_{22}	7,200	5,400



Map of $x_2(L,t)$ at t=0



Map of $x_2(L,t)$ at t=15

Figure 35. Initial and Final Contour Maps

Since the initial a_{ij} are identical to those given in Table 7 (page 112), the initial $x_2(L,t)$ contour map is identical to that given in Figure 28 (page 114). This initial map, and also the final map, are given in Figure 35. In comparing the two maps of Figure 35, it may be seen that the optimum drifted from the upper right corner, at t=0, to the top of the curved ridge, at t=15.

Given in Figure 36 are the $x_2(L,t)$ resulting from the efforts of the multiplexed gradient, one-at-a-time, and standard gradient techniques to track the drifting optimum.

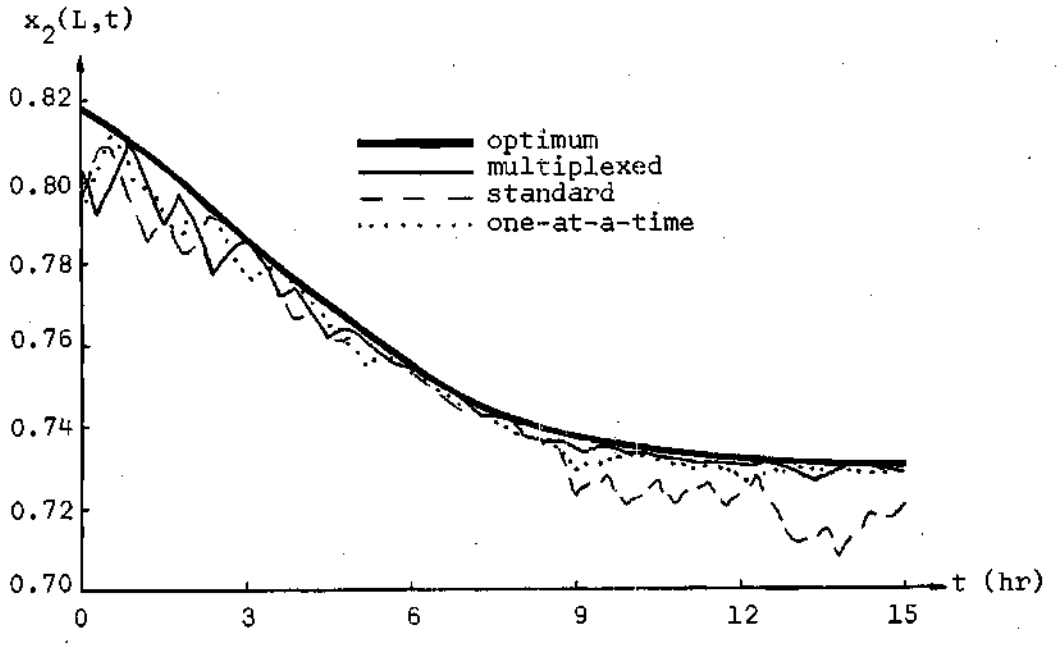


Figure 36. Response to a Drifting Optimum

Each was started at t=0 at set points of $u_1(t) = u_2(t) = 395°K$. For purposes of comparison, the $x_2(L,t)$ at the drifting optimum is also shown.

For $t \leq 9$ the optimum was constrained by the boundaries and all three techniques were equally capable of maintaining a limit cycle about the optimum. For $t>9$, however, the optimum was at the top of the drifting curved ridge and the multiplexed gradient and one-at-a-time techniques were visibly superior to the standard gradient technique (see Figure 36). The reason for this may be deduced from (4-6). The multiplexed gradient and one-at-a-time techniques establish locally-linear system models in roughly one system residence time, so that for them the factor $(t^a - t^{a-1})$ is L/v. The standard gradient technique, however, requires roughly two (in general, n) residence times to establish the locally-linear model, and hence for it the factor $(t^a - t^{a-1})$ is 2 L/v. It follows that the standard gradient technique, to satisfy (4-6), must operate with twice as large a $\partial \underline{H}/\partial u_{ja}\big|_{a-1}$ as that required by the other two techniques. In short, it must operate further from the optimum where the $\partial \underline{H}/\partial u_{ja}\big|_{a-1}$ are larger.

To obtain a quantitative measure of the comparative tracking abilities, the integral

$$\int_0^{15} [x_2(L,t)_{\text{optimum}} - x_2(L,t)_{\text{technique}}]dt \qquad (8\text{-}32)$$

was evaluated for each of the three techniques shown in Figure 36. The resulting values, proportional to the dollar value of the yield not achieved, were

Multiplexed gradient:	0.037
One-at-a-time:	0.040
Standard gradient:	0.103.

(8-33)

The superiority of the multiplexed gradient and one-at-a-time techniques is evident.

Noise-Corrupted State-Constrained System

To compare the ability of the three techniques to function in a noise-corrupted environment, the system of Figure 27 (page 112) was modified as shown in Figure 37. The χ_{jk} were independent samples of a random process uniform on [-0.5,+0.5] and the χ were independent samples of a Gaussian random process with mean zero and standard deviation σ.

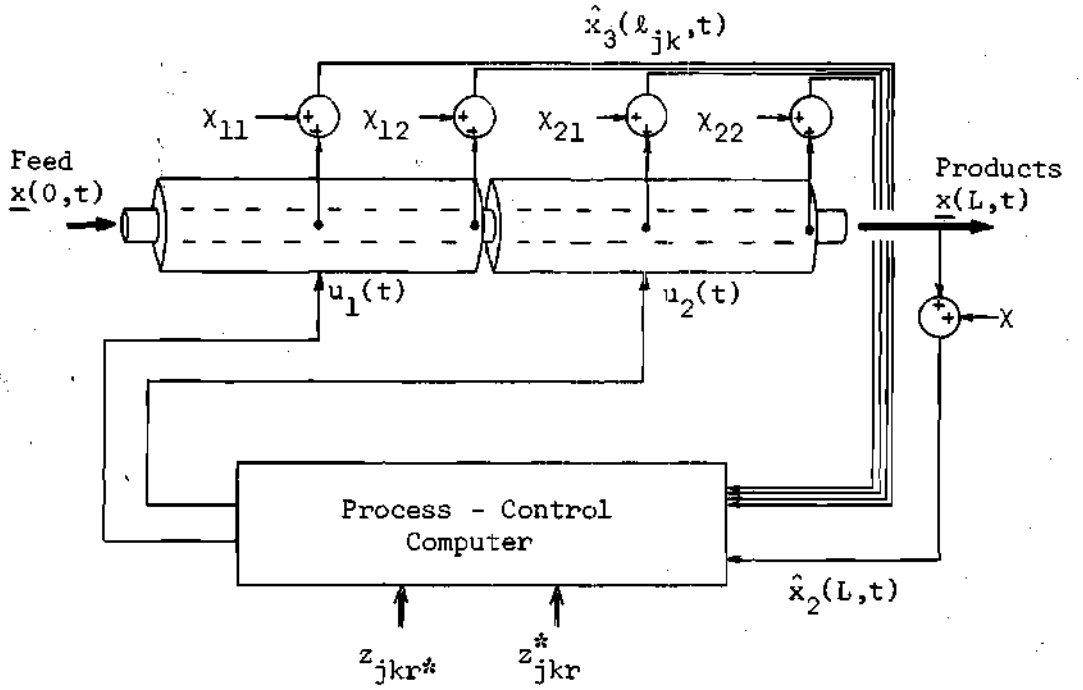


Figure 37. Noise-Corrupted State-Constrained System

Simulation runs corresponding to those which produced Figure 30 (page 116) were made with several values of σ. Given in Figure 38 are

the resulting $x_2(L,t)$ at t=12. This time was chosen for illustration because all of the noise-free response curves (Figure 30, page 116) were still rising at that time. Insofar as possible, the scales of Figure 38 correspond to those of Figure 17 (page 81). The difference $[x_2(L,12)-x_2(L,0)]$ is proportional to $E\{\Delta F\}$, and the abscissa scale factor 0.015 is an approximation to the steady-state $\frac{\partial x_2(L,t)}{\partial u_1(t)} \Delta u_1$ and $\frac{\partial x_2(L,t)}{\partial u_2(t)} \Delta u_2$ as measured on Figure 28 (page 114).

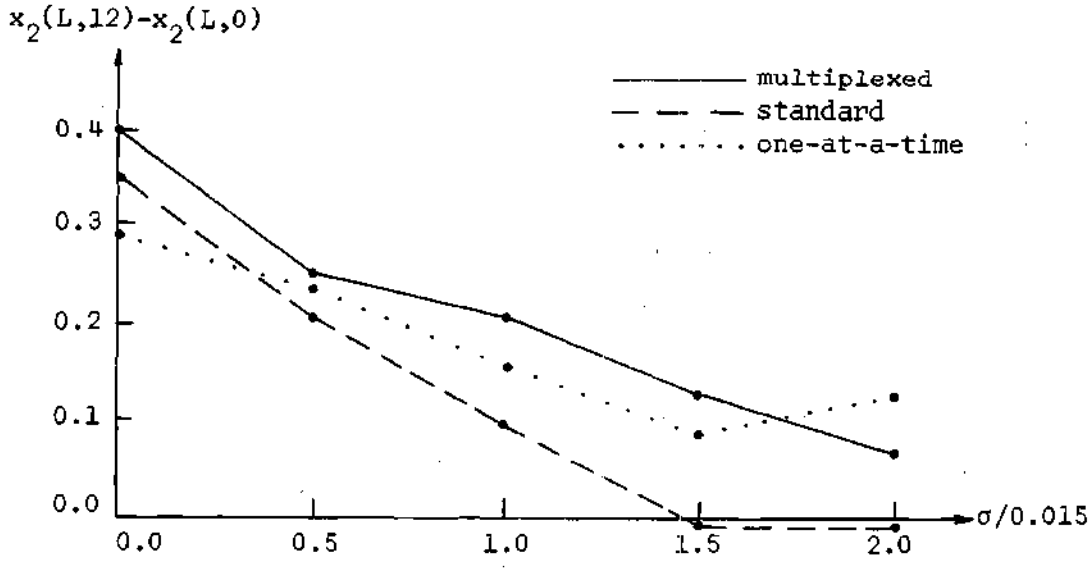


Figure 38. Yield as a Function of Noise Standard Deviation

Two observations may be made from Figure 38. First, the multiplexed gradient curve is seen to follow the n≥2 curve of Figure 17 (page 81) rather well, considering that the curve of Figure 38 includes the effects of state constraints. Second, the superiority of the multiplexed gradient and one-at-a-time techniques is evident, with the multiplexed gradient in general the better of the two. This is not

surprising, for the multiplexed gradient technique operates on all of the available data at each algorithmic cycle while the other two techniques, during their one-dimensional searches, operate on only part of the available data. In addition, the standard gradient technique requires a sequence of n+1 = 3 reasonably accurate $\hat{x}_2(L,t)$ for determination of the gradient direction, a sequence whose probability of occurrence rapidly approaches zero as σ increases.

Finally, it should be remarked that in all of the multiplexed gradient simulation runs ($\sigma/0.015 = 0.0,0.5,1.0,1.5,2.0$; $0 \le t \le 18$), the maximum $x_3(\ell_{jk},t)$ encountered was 400.66°K, excellent satisfaction of the 400.00°K constraint considering that the process-control computer was given $x_3(\ell_{jk},t)$ corrupted by noise uniform on [-0.50,+0.50].

Restrictions on Use of Algorithm

It seems appropriate, during this presentation of example systems, to consider the restrictions on the use of the algorithm. Most of these are self-evident when it is recalled that the algorithm is based on the system model of (3-50)-(3-51) and the local linearizations of (4-7) and (5-10). But specifically, fluid velocities and system settling times, while not required to be constant, are required to lie within the bounds assumed. Set point controllers, while not required to be free of sustained oscillations, are required to satisfy the assumption of (3-11). Boundary conditions, while not required to be time-invariant, are required to be free of abrupt changes. And system parameters in general, while not required to be constant, are required to vary slowly enough that (4-6) and (5-9) are satisfied,

except for the former at the optimum. Finally, measurement noise, while not required to be negligible, is required to be small enough that reasonably accurate constraint models can be obtained, and that (6-11) is satisfied except at the optimum.

Several remarks tempering these admonitions are in order. First, it should be noted that the continual updating of the locally-linear system model permits rather severe nonlinearities to be handled, such as the concave response surface of Figure 22 (page 101) and the multimodal curved-ridge response surface of Figure 25 (page 109). Multimodal response surfaces, of course, require that the initial set points be such that the local optimum achieved is the global one. Nonlinear constraint boundaries may also be handled, though the radius of curvature of such boundaries must be on the order of several Δu_j, as in Figure 33 (page 120).

Second, it should be noted that abrupt changes in system boundary conditions create no problems in the unconstrained and control-constrained cases. Set point adjustments may not increase yield for one, or at most two, algorithmic cycles, but the algorithm is self-recovering. The same is true in the state-constrained case, except that there may be temporary constraint violations until the new boundaries are located and the set points are adjusted to satisfy them. An alternative to permitting such violations is to apply methods of transient control[18,52] to establish new set points, after which the algorithm may be restarted.

Finally, it should be noted that the Δu_j may generally be chosen large enough to insure satisfaction of (4-6) and (5-9), excepting the

former at the optimum, as was done for the drift of (8-30). Should this not be possible, the basic algorithm may be extended to encompass the time-dependent locally-linear models of (4-5) and (5-8).[53]

Summary of Comparative Studies

The simulation studies of this chapter provide a basis for a summary comparison of the multiplexed gradient, one-at-a-time, and standard gradient techniques. Specifically, with respect to the four criteria given at the beginning of the chapter, the following judgments may be made.

(a) *Speed in Achieving the Optimum.* The multiplexed gradient technique is the fastest, with the standard gradient technique a close second and the one-at-a-time technique a poor third. The one-at-a-time technique is often slower by a factor of two or more.

(b) *Ability to Converge to a Constrained Optimum.* In general, all three techniques are equally good. However, the standard gradient technique may fail on curved ridges (Figure 25, page 109) and the one-at-a-time technique will fail with certain constraint geometries (Figure 19, page 90).

(c) *Ability to Track a Drifting Optimum.* The multiplexed gradient and one-at-a-time techniques are equally good, with the standard gradient technique a poor third. Yield loss with the standard gradient technique may be greater by a factor of two or more.

(d) *Ability to Function in a Noise-Corrupted Environment.* The multiplexed gradient and one-at-a-time techniques are equally good,

with the standard gradient technique a poor third. The standard gradient technique fails completely at high noise levels.

Thus, the multiplexed gradient technique combines the best qualities of the other two techniques. It has the speed of the standard gradient technique and the drift and noise insensitivity of the one-at-a-time technique. And from Chapter VII, its hardware requirements are equivalent to, and in some instances less than, those of the other two techniques.

CHAPTER IX

CONCLUSIONS AND RECOMMENDATIONS

This dissertation has outlined the development of an extremum control algorithm, suitable for implementation on a process-control computer, for a class of nonlinear time-varying distributed-parameter processes with multiple inputs and multiple outputs.

Conclusions

There are two principal contributions of the research. The first is the unification of a number of diverse system types and control objectives into a common mathematical framework. The second is the development of a control algorithm which uses to advantage the transport delays inherent in these systems.

More specifically, a class of systems has been defined, and a control problem specified, which embraces the extremum control requirements of multistage tubular and stirred-tank reactor systems. Significant in the problem definition is the inclusion of transport delay between stages and the consideration of set point controller dynamics, control and state constraints, and measuring instrument time delay. Particularly significant is the assumption, made at the outset, that system states are constrained only at discrete spatial points. The problem definition, by virtue of its generality and practicality, permits results to be applied to a wide variety of system configurations.

A general system model, suitable for use in the development of the control algorithm, has been developed from the ordinary and partial differential equations describing the class of systems considered. Significant in this model is the unification of lumped-parameter (stirred-tank) and distributed-parameter (tubular) systems into a common mathematical framework.

An extremum control algorithm has been developed which utilizes system transport delays to interlace the effects of several set point combinations. This interlacing is controlled by equations specifying the times at which controller set points are to be changed and system measurements are to be made. Convergence and closed-loop stability of the algorithm have been established.

The algorithm has been extended to handle system control and state constraints. Procedures have been developed for avoiding division and square root operations in the computation of the gradient projection matrix. Convergence of the control-constrained and state-constrained algorithms has been established.

The algorithm has been shown to be relatively insensitive to the effects of additive white Gaussian measurement noise.

The hardware required to implement the algorithm has been established and compared to that required to implement two currently available procedures. Requirements have been shown to be equivalent to, and in some instances less than, those of the currently available procedures.

The operational characteristics of the algorithm have been

investigated via digital simulation of multistage tubular and stirred-tank reactor systems. Through these studies the algorithm has been shown to have excellent speed, convergence, tracking, and noise insensitivity properties.

Recommendations

A logical extension of the thesis research is the consideration of systems described by the partial differential equation

$$\frac{\partial \underline{x}(\ell,t)}{\partial t} + v(\ell,t)\frac{\partial \underline{x}(\ell,t)}{\partial \ell} + d(\ell,t)\frac{\partial^2 \underline{x}(\ell,t)}{\partial \ell^2} = \underline{f}[\underline{x}(\ell,t),u(\ell,t),\ell,t]. \quad (9\text{-}1)$$

This equation is more general than (2-1) and may be used to describe the behavior of tubular reactors with axial diffusion.[52,59,62]

Procedures for establishing stability[54-56] and optimal steady-state conditions[57-61] in systems described by (9-1) are rather well developed, but procedures for dynamic control are few.[52,62] It should not be overly difficult to extend the analysis of Chapter III to the case of (9-1), using the method of regular perturbation[63-66] as a tool for developing a suitable system model. This model should be of the same form as (3-50) and (3-51). Only the definitions corresponding to (B-10)-(B-13) should be different.

The results of Chapters IV-VII should be directly applicable to the diffusion case. No further work should be necessary here.

Verification of results would not be as simple as for the case considered in this research, for (9-1) does not admit transformation by

the method of characteristics. Probably the best route to simulation would be the classical continuous-time discrete-space approach,[52,67] with discrete points including those points at which system states would be monitored by measuring instruments. Alternatively, results might be verified by experimental means. In either case, it would be appropriate to investigate the closed-loop stability of the control algorithm when coupled to a system described by (9-1).

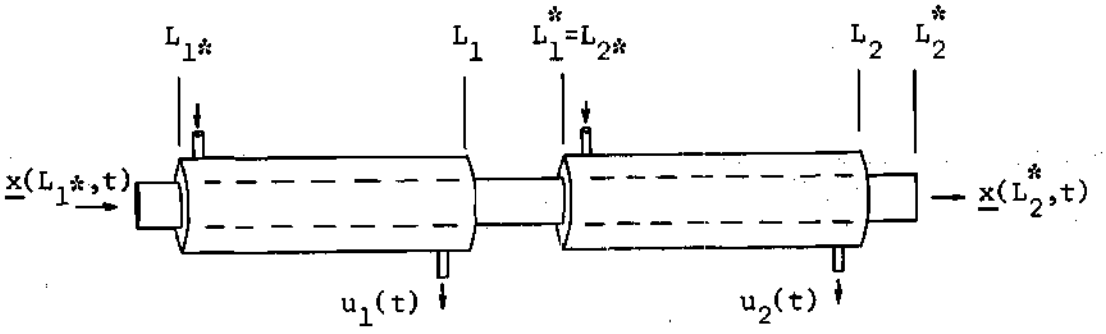
APPENDICES

APPENDIX A

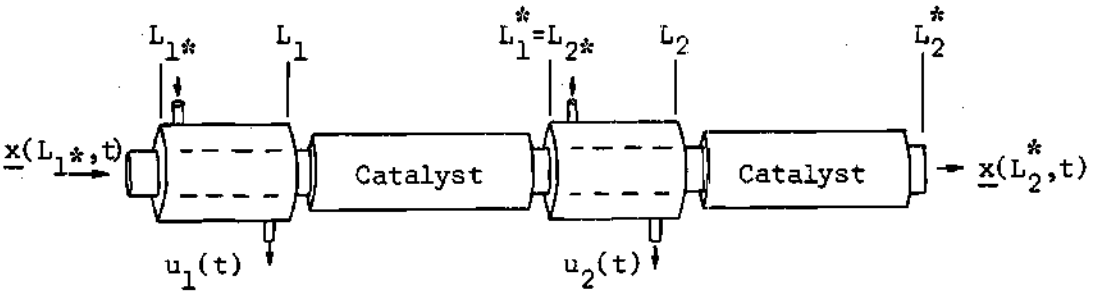
SCHEMATIC DIAGRAMS OF REPRESENTATIVE MULTISTAGE SYSTEMS

Given below are schematic diagrams of several systems embraced by the formulation of Chapter II. In the systems shown, the controls $u_1(t)$ and $u_2(t)$ are the temperatures of the coolants, or heating media, if heating is required, and the states $\underline{x}(\ell,t)$ are the concentrations and temperature of the reacting species.

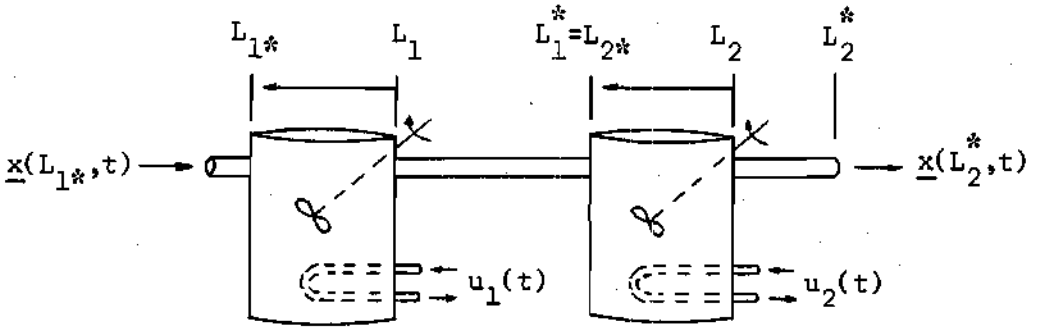
Multijacketed tubular reactor:



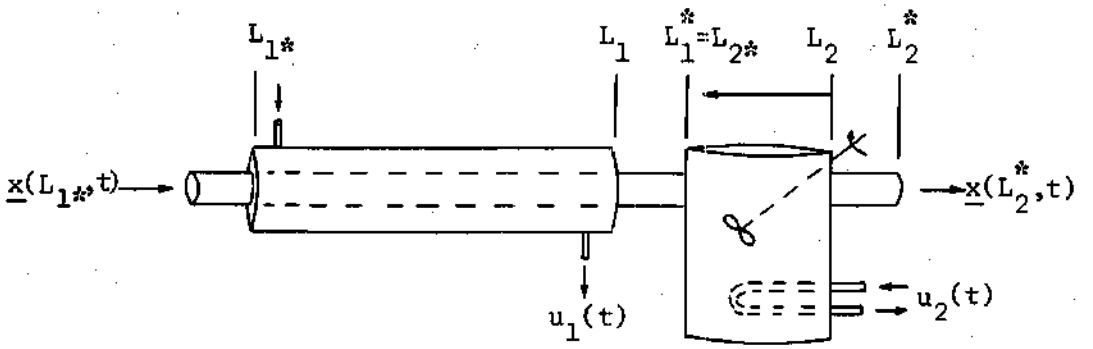
Sequence of plug-flow adiabatic catalytic reactors:



Sequence of stirred-tank reactors:



Series combination of tubular and stirred-tank reactors:



APPENDIX B

DEVIATION OF GENERAL JTH STAGE MODEL

Given below is the derivation of the general jth stage model used to generate the complete system model. First, to establish (3-47), define the function $\underline{X}_j[\ell;\underline{x},u,t]$ by

$$\underline{X}_j[\ell;\underline{x},u,t] = \begin{cases} \underline{S}_j[\ell;\underline{x},u,t] & \text{basic,} \quad L_{j*}<\ell\leq L_j \\ \underline{W}_j[\ell;\underline{x},u,t] & \text{basic,} \quad L_j<\ell\leq L_j^* \\ \underline{S}_j''[\underline{x},u,t] & \text{degenerate,} \quad \ell=L_j \\ \underline{W}_j'[\ell;\underline{x},u,t] & \text{degenerate,} \quad L_j<\ell\leq L_j^* \end{cases} \tag{B-1}$$

and the function $\bar{\sigma}_j(\ell,t)$ by

$$\bar{\sigma}_j(\ell,t) = \begin{cases} \sigma_j(\ell,t) & \text{basic,} \quad L_{j*}<\ell\leq L_j \\ \sigma_j(\ell,t) & \text{basic,} \quad L_j<\ell\leq L_j^* \\ 0 & \text{degenerate,} \quad \ell=L_j \\ \rho_j(\ell,t) & \text{degenerate,} \quad L_j<\ell\leq L_j^*. \end{cases} \tag{B-2}$$

It then follows that (3-13), (3-29), (3-39), and (3-44) may be represented by the single equation

$$\underline{x}(\ell,t) = \underline{X}_j\left[\ell;\underline{x}\left(L_{j*},t-\bar{\sigma}_j(\ell,t)\right),u_{ja},t\right] \quad L_{j*}\leq\ell\leq L_j^* \tag{B-3}$$

given in Chapter III as (3-47).

Next, to establish (3-48), use (3-49) to obtain upper and lower bounds on the solutions to (3-1) and (3-2). It then follows from (3-8) and (3-22) that

$$\int_{L_{j*}}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda \le \sigma_j(\ell,t) \le \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda \tag{B-4}$$

$$\int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda \le \rho_j(\ell,t) \le \int_{L_j}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda. \tag{B-5}$$

The four validity conditions (3-20), (3-32), (3-40), and (3-46) may then be replaced by the more restrictive conditions

$$\begin{aligned} \text{Max}\{t_{ja} + \tau_{jc} + \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda,\ s_{j*} + \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda\} \\ \le t \le \text{Min}\{t_{j,a+1},\ s_j^* + \int_{L_{j*}}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda\} \end{aligned} \tag{B-6}$$

$$\begin{aligned} \text{Max}\{t_{ja} + \tau_{jc} + \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda,\ s_{j*} + \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda\} \\ \le t \le \text{Min}\{t_{j,a+1} + \int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda,\ s_j^* + \int_{L_{j*}}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda\} \end{aligned} \tag{B-7}$$

$$\text{Max}\{t_{ja} + \tau_{jc} + \tau_{js},\ s_{j*} + \tau_{js}\} \le t \le \text{Min}\{t_{j,a+1},\ s_j^*\} \tag{B-8}$$

$$\text{Max}\{t_{ja} + \tau_{jc} + \tau_{js} + \int_{L_j}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda,\ s_{j*} + \tau_{js} + \int_{L_j}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda\} \quad \text{(B-9)}$$

$$\le t \le \text{Min}\{t_{j,a+1} + \int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda,\ s_j^* + \int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda\}.$$

Next, define the functions $\rho_j^*(\ell)$, $\sigma_j^*(\ell)$, $\rho_{j*}(\ell)$, and $\sigma_{j*}(\ell)$ by

$$\rho_j^*(\ell) = \begin{cases} \tau_{jc} + \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda & \text{basic}, \quad L_{j*} < \ell \le L_j \\ \tau_{jc} + \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda & \text{basic}, \quad L_j < \ell \le L_j^* \\ \tau_{jc} + \tau_{js} & \text{degenerate}, \quad \ell = L_j \\ \tau_{jc} + \tau_{js} + \int_{L_j}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda & \text{degenerate}, \quad L_j < \ell \le L_j^* \end{cases} \quad \text{(B-10)}$$

$$\sigma_j^*(\ell) = \begin{cases} \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda & \text{basic}, \quad L_{j*} < \ell \le L_j \\ \int_{L_{j*}}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda & \text{basic}, \quad L_j < \ell \le L_j^* \\ \tau_{js} & \text{degenerate}, \quad \ell = L_j \\ \tau_{js} + \int_{L_j}^{\ell} \frac{1}{v_*(\lambda)}\, d\lambda & \text{degenerate}, \quad L_j < \ell \le L_j^* \end{cases} \quad \text{(B-11)}$$

$$\rho_{j*}(\ell) = \begin{cases} 0 & \text{basic,} \quad L_{j*} < \ell \le L_j \\ \int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda & \text{basic,} \quad L_j < \ell \le L_j^* \\ 0 & \text{degenerate,} \quad \ell = L_j \\ \int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda & \text{degenerate,} \quad L_j < \ell \le L_j^* \end{cases} \tag{B-12}$$

$$\sigma_{j*}(\ell) = \begin{cases} \int_{L_{j*}}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda & \text{basic,} \quad L_{j*} < \ell \le L_j \\ \int_{L_{j*}}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda & \text{basic,} \quad L_j < \ell \le L_j^* \\ 0 & \text{degenerate,} \quad \ell = L_j \\ \int_{L_j}^{\ell} \frac{1}{v^*(\lambda)}\, d\lambda & \text{degenerate,} \quad L_j < \ell \le L_j^*. \end{cases} \tag{B-13}$$

It then follows that (B-6), (B-7), (B-8), and (B-9) may be represented by the single inequality

$$\text{Max}\{t_{ja} + \rho_j^*(\ell),\ s_{j*} + \sigma_j^*(\ell)\} \tag{B-14}$$

$$\le t \le \text{Min}\{t_{j,a+1} + \rho_{j*}(\ell),\ s_j^* + \sigma_{j*}(\ell)\}$$

given in Chapter III as (3-48).

APPENDIX C

DERIVATION OF COMPLETE SYSTEM MODEL

The complete system model may be obtained by coupling the general jth stage model to the corresponding (j-1)st stage model, and continuing such coupling until the 1st stage is included.

Model of (j-1)st Stage

To begin this coupling, use (2-10) to obtain

$$L^*_{j-1} = L_{j*}. \tag{C-1}$$

It follows from (C-1) that the states $\underline{x}\left(L_{j*}, t-\bar{\sigma}_j(\ell,t)\right)$ found in (3-47) are identically equal to the states $\underline{x}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)$. The latter may be modeled by (3-47) if j is replaced by j-1, ℓ is replaced by L^*_{j-1}, and t is replaced by $t-\bar{\sigma}_j(\ell,t)$:

$$\underline{x}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right) = \underline{X}_{j-1}\Big[L^*_{j-1}; \underline{x}\Big(L_{j-1*}, t-\bar{\sigma}_j(\ell,t) - \bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\Big), u_{j-1,a}, t-\bar{\sigma}_j(\ell,t)\Big]. \tag{C-2}$$

The s_{j*} and s^*_j found in (3-48) are the bounds on the time interval in which the states $\underline{x}(L_{j*},t)$ are at steady state. In view of (C-1), these bounds are the same as those imposed on $\underline{x}(L^*_{j-1},t)$. It follows

that s_{j*} and s_j^* may be obtained from (3-48) by setting j to j-1 and ℓ to L_{j-1}^*:

$$s_{j*} = \text{Max}\{t_{j-1,a} + \rho_{j-1}^*(L_{j-1}^*),\ s_{j-1*} + \sigma_{j-1}^*(L_{j-1}^*)\} \tag{C-3}$$

$$s_j^* = \text{Min}\{t_{j-1,a+1} + \rho_{j-1*}(L_{j-1}^*),\ s_{j-1}^* + \sigma_{j-1*}(L_{j-1}^*)\}. \tag{C-4}$$

Coupling the jth and (j-1)st Stages

Inserting (C-2) into (3-47), the coupled solution equation

$$\underline{x}(\ell,t) = \underline{X}_j\Big[\ell;\underline{X}_{j-1}\Big[L_{j-1}^*;\underline{x}\Big(L_{j-1*},t-\bar{\sigma}_j(\ell,t) - \bar{\sigma}_{j-1}\big(L_{j-1}^*,t-\bar{\sigma}_j(\ell,t)\big)\Big),u_{j-1,a},t-\bar{\sigma}_j(\ell,t)\Big],u_{ja},t\Big] \tag{C-5}$$

$$L_{j*}<\ell\leq L_j^*$$

is obtained. Similarly, inserting (C-3) and (C-4) into (3-48), the validity condition associated with (C-5) is obtained:

$$\text{Max}\{t_{ja} + \rho_j^*(\ell),\ t_{j-1,a} + \rho_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell), \tag{C-6}$$

$$s_{j-1*} + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell)\} \leq t \leq$$

$$\text{Min}\{t_{j,a+1} + \rho_{j*}(\ell),\ t_{j-1,a+1} + \rho_{j-1*}(L_{j-1}^*) + \sigma_{j*}(\ell),$$

$$s_{j-1}^* + \sigma_{j-1*}(L_{j-1}^*) + \sigma_{j*}(\ell)\}.$$

Model of (j-2)nd Stage

The model required for insertion into (C-5) may be obtained from (3-47) by setting j to j-2, ℓ to L^*_{j-2}, and t to $t - \bar{\sigma}_j(\ell,t) - \bar{\sigma}_{j-1}\left(L^*_{j-1}, t - \bar{\sigma}_j(\ell,t)\right)$:

$$\underline{x}\left(L^*_{j-2}, t-\bar{\sigma}_j(\ell,t)-\bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\right) \tag{C-7}$$

$$= \underline{X}_{j-2}\left[L^*_{j-2}; \underline{x}\left(L_{j-2^*}, t-\bar{\sigma}_j(\ell,t)-\bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\right.\right.$$

$$\left.\left. - \bar{\sigma}_{j-2}\left(L^*_{j-2}, t-\bar{\sigma}_j(\ell,t)-\bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\right)\right), u_{j-2,a},\right.$$

$$\left. t - \bar{\sigma}_j(\ell,t) - \bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\right].$$

Similarly, the s_{j-1^*} and s^*_{j-1} required for insertion into (C-6) may be obtained from (3-48) by setting j to j-2 and ℓ to L^*_{j-2}:

$$s_{j-1^*} = \text{Max}\{t_{j-2,a} + \rho^*_{j-2}(L^*_{j-2}),\ s_{j-2^*} + \sigma^*_{j-2}(L^*_{j-2})\} \tag{C-8}$$

$$s^*_{j-1} = \text{Min}\{t_{j-2,a+1} + \rho_{j-2^*}(L^*_{j-2}),\ s^*_{j-2} + \sigma_{j-2^*}(L^*_{j-2})\}. \tag{C-9}$$

Coupling the jth, (j-1)st, and (j-2)nd Stages

Inserting (C-7) into (C-5), the twice coupled solution equation

$$\underline{x}(\ell,t) = \underline{X}_j\left[\ell; \underline{X}_{j-1}\left[L^*_{j-1}; \underline{X}_{j-2}\left[L^*_{j-2}; \underline{x}\left(L^*_{j-2^*}, t-\bar{\sigma}_j(\ell,t)\right.\right.\right.\right. \tag{C-10}$$

$$- \bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right) - \bar{\sigma}_{j-2}\Big[L^*_{j-2}, t-\bar{\sigma}_j(\ell,t)-\bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\Big], u_{j-2,a}, t-\bar{\sigma}_j(\ell,t)-\bar{\sigma}_{j-1}\left(L^*_{j-1}, t-\bar{\sigma}_j(\ell,t)\right)\Big], u_{j-1,a}, t-\bar{\sigma}_j(\ell,t)\Big], u_{ja}, t\Big] \qquad L_{j*}<\ell\leq L^*_j$$

is obtained. Similarly, inserting (C-8) and (C-9) into (C-6), the validity condition associated with (C-10) is obtained:

$$\begin{aligned} \text{Max}\{&t_{ja} + \rho^*_j(\ell),\ t_{j-1,a} + \rho^*_{j-1}(L^*_{j-1}) + \sigma^*_j(\ell), \\ &t_{j-2,a} + \rho^*_{j-2}(L^*_{j-2}) + \sigma^*_{j-1}(L^*_{j-1}) + \sigma^*_j(\ell), \\ &s_{j-2*} + \sigma^*_{j-2}(L^*_{j-2}) + \sigma^*_{j-1}(L^*_{j-1}) + \sigma^*_j(\ell)\} \leq t \leq \\ \text{Min}\{&t_{j,a+1} + \rho_{j*}(\ell),\ t_{j-1,a+1} + \rho_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell), \\ &t_{j-2,a+1} + \rho_{j-2*}(L^*_{j-2}) + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell), \\ &s^*_{j-2} + \sigma_{j-2*}(L^*_{j-2}) + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell)\}. \end{aligned} \tag{C-11}$$

Complete System Model

The complete system model may be obtained by continuing this coupling until the 1st stage is included. The right-hand side of the

equation corresponding to (C-10) may be envisioned as a complex expression involving the functions $\underline{X}_j, \underline{X}_{j-1}, \ldots, \underline{X}_1$, the independent variable ℓ, the 1st stage inlet states $\underline{x}(L_{1*}, t)$, and the parameters $u_{ja}, u_{j-1,a}, \ldots, u_{1a}$ and t. This expression may be simplified by defining a new function $\underline{Y}_j$ whose form permits the equation to be written

$$\underline{x}(\ell,t) = \underline{Y}_j(\ell; u_{1a}, \ldots, u_{ja}, t) \qquad L_{j*} < \ell \leq L_j^*. \tag{C-12}$$

The dependence of $\underline{Y}_j$ on t embraces the functional dependence of $\underline{x}(\ell,t)$ on the 1st stage inlet states which, from (2-8) and (2-6), are the system boundary conditions $\underline{x}_B(t)$. Equation (C-12) is given in Chapter III as (3-50).

The validity condition associated with (C-12) may be obtained by replacing the s_{j-2*} and s_{j-2}^* of (C-11) with expressions analogous to (C-8) and (C-9), and continuing such replacements until only s_{1*} and s_1^* remain:

$$\begin{aligned} \text{Max}\{&t_{ja} + \rho_j^*(\ell),\ t_{j-1,a} + \rho_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell), \\ &t_{j-2,a} + \rho_{j-2}^*(L_{j-2}^*) + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell), \ldots, \\ &t_{1a} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell), \\ &s_{1*} + \sigma_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell)\} \leq t \leq \end{aligned} \tag{C-13}$$

$$\text{Min}\{t_{j,a+1} + \rho_{j*}(\ell),\ t_{j-1,a+1} + \rho_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell),$$

$$t_{j-2,a+1} + \rho_{j-2*}(L^*_{j-2}) + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell),\ldots,$$

$$t_{1,a+1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell),$$

$$s^*_1 + \sigma_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{j-1*}(L^*_{j-1}) + \sigma_{j*}(\ell)\}.$$

The dependence of this inequality on s_{1*} and s^*_1 may be removed by assuming that the boundary conditions $\underline{x}_B(t)$ are at steady state for all $t \geq 0$. The resulting validity condition is given in Chapter III as (3-51).

APPENDIX D

DERIVATION OF SEQUENCING EQUATIONS

This appendix contains the derivation of the equations establishing the times at which system set points are changed and the times at which system measurements are made. These equations are derived by combining and simplifying a number of validity conditions and inequality constraints, and then transforming these inequalities into equalities by requiring that each algorithmic cycle be completed in the minimum possible time.

Measurement Validity Conditions

Consider first the validity conditions imposed on the measurements (4-17). These conditions are obtained from (3-51) by setting j to n, ℓ to L, t to the required $t^{j,a-1}$, and "a" to the index corresponding to the set points in the required measurement:

$$\text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \qquad \text{(D-1)}$$

$$t_{n-2,a-1} + \rho_{n-2}^*(L_{n-2}^*) + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L),\ \ldots,$$

$$t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\} \leq t^{0,a-1} \leq$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),\ t_{n-1,a} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{n-2,a} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),\ldots,$$

$$t_{1a} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}$$

$$\text{Max}\{t_{n,a-1} + \rho^*_n(L),\ t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),\ldots, \qquad \text{(D-2)}$$

$$t_{2,a-1} + \rho^*_2(L^*_2) + \sigma^*_3(L^*_3) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)\} \le t^{1,a-1} \le$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),\ t_{n-1,a} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),\ldots,$$

$$t_{2a} + \rho_{2*}(L^*_2) + \sigma_{3*}(L^*_3) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),$$

$$t_{1,a-1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}$$

$$\text{Max}\{t_{n,a-1} + \rho^*_n(L),\ t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),\ldots, \qquad \text{(D-3)}$$

$$t_{3,a-1} + \rho^*_3(L^*_3) + \sigma^*_4(L^*_4) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),$$

$$t_{2,a-2} + \rho^*_2(L^*_2) + \sigma^*_3(L^*_3) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)\} \le t^{2,a-1} \le$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),\ t_{n-1,a} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),\ldots,$$

$$t_{3a} + \rho_{3*}(L^*_3) + \sigma_{4*}(L^*_4) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),$$

$$t_{2,a-1} + \rho_{2*}(L^*_2) + \sigma_{3*}(L^*_3) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),$$

$$t_{1,a-1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}$$

$$\vdots$$

$$\text{Max}\{t_{n,a-1} + \rho^*_n(L),\ t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L), \qquad \text{(D-4)}$$

$$t_{n-2,a-2} + \rho^*_{n-2}(L^*_{n-2}) + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),\ldots,$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)\} \leq t^{n-2,a-1} \leq$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),\ t_{n-1,a} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),$$

$$t_{n-2,a-1} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L),\ldots,$$

$$t_{1,a-1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}$$

$$\text{Max}\{t_{n,a-1} + \rho^*_n(L),\ t_{n-1,a-2} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L), \qquad \text{(D-5)}$$

$$t_{n-2,a-2} + \rho^*_{n-2}(L^*_{n-2}) + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L),\ldots,$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\} \leq t^{n-1,a-1} \leq$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),\ t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L), \ldots,$$

$$t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)\}$$

$$\text{Max}\{t_{n,a-2} + \rho_n^*(L),\ t_{n-1,a-2} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \qquad \text{(D-6)}$$

$$t_{n-2,a-2} + \rho_{n-2}^*(L_{n-2}^*) + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \ldots,$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\} \leq t^{n,a-1} \leq$$

$$\text{Min}\{t_{n,a-1} + \rho_{n*}(L_n^*),\ t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L), \ldots,$$

$$t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)\}.$$

Inequalities from Validity Conditions

Now consider a set of inequalities which may be extracted from (D-2)-(D-5). From (D-2),

$$t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L) \qquad \text{(D-7)}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \ldots,$$

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\}.$$

From (D-3),

$$t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L) \qquad \text{(D-8)}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \ldots,$$

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\}.$$

From (D-4),

$$t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L) \qquad \text{(D-9)}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\}.$$

And from (D-5),

$$t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L) \geq t_{n,a-1} + \rho_n^*(L). \qquad \text{(D-10)}$$

To put (D-7)-(D-10) in a more useful form, note from (B-10)-(B-13) that

$$\rho_j^*(\ell) \geq \rho_{j*}(\ell) \tag{D-11}$$

and

$$\sigma_j^*(\ell) \geq \sigma_{j*}(\ell). \tag{D-12}$$

Using (D-11) and (D-12), it follows from (D-7) that

$$t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \tag{D-13}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \ldots,$$

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\}.$$

Similarly, from (D-8),

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \tag{D-14}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \ldots,$$

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\}.$$

From (D-9),

$$t_{n-2,a-1} + \rho_{n-2}^*(L_{n-2}^*) + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \tag{D-15}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_n^*(L),\ t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\}.$$

And finally, from (D-10),

$$t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \geq t_{n,a-1} + \rho^*_n(L). \qquad \text{(D-16)}$$

Now reversing the application of (D-11) and (D-12), it follows from (D-7) that

$$t_{1,a-1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L) \qquad \text{(D-17)}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_{n*}(L),\ t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L), \ldots,$$

$$t_{2,a-1} + \rho_{2*}(L^*_2) + \sigma_{3*}(L^*_3) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}.$$

Similarly, from (D-8),

$$t_{2,a-1} + \rho_{2*}(L^*_2) + \sigma_{3*}(L^*_3) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L) \qquad \text{(D-18)}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_{n*}(L),\ t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L), \ldots,$$

$$t_{3,a-1} + \rho_{3*}(L^*_3) + \sigma_{4*}(L^*_4) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}.$$

From (D-9),

$$t_{n-2,a-1} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L) \qquad \text{(D-19)}$$

$$\geq \text{Max}\{t_{n,a-1} + \rho_{n*}(L),\ t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}.$$

And finally, from (D-10),

$$t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L) \geq t_{n,a-1} + \rho_{n*}(L). \qquad \text{(D-20)}$$

Simplified Validity Conditions

Equations (D-13)-(D-16) may be used to simplify the Max{·} of (D-1)-(D-6). Similarly, (D-17)-(D-20) may be used to simplify the Min{·} of (D-1)-(D-6). The results, respectively, are

$$t_{1,a-1} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \qquad \text{(D-21)}$$

$$\leq t^{0,a-1} \leq t_{na} + \rho_{n*}(L)$$

$$\text{Max}\{t_{2,a-1} + \rho^*_2(L^*_2) + \sigma^*_3(L^*_3) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L), \qquad \text{(D-22)}$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)\} \leq t^{1,a-1} \leq$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),$$

$$t_{1,a-1} + \rho_{1*}(L^*_1) + \sigma_{2*}(L^*_2) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)\}$$

$$\text{Max}\{t_{3,a-1} + \rho^*_3(L^*_3) + \sigma^*_4(L^*_4) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L), \qquad \text{(D-23)}$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)\} \leq t^{2,a-1} \leq$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),$$

$$t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)\}$$

$$\vdots$$

$$\text{Max}\{t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \qquad \text{(D-24)}$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\} \le t^{n-2,a-1} \le$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),$$

$$t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)\}$$

$$\text{Max}\{t_{n,a-1} + \rho_n^*(L), \qquad \text{(D-25)}$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)\} \le t^{n-1,a-1} \le$$

$$\text{Min}\{t_{na} + \rho_{n*}(L),$$

$$t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)\}$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-26)}$$

$$\le t^{n,a-1} \le t_{n,a-1} + \rho_{n*}(L).$$

Inequalities from Simplified Validity Conditions

Now consider a set of inequalities which may be extracted from (D-21)-(D-26). First, from (D-21)-(D-24), respectively,

$$t_{na} + \rho_{n*}(L) \tag{D-27}$$

$$\geq t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{na} + \rho_{n*}(L) \tag{D-28}$$

$$\geq t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{na} + \rho_{n*}(L) \tag{D-29}$$

$$\geq t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{3,a-1} + \rho_{3*}(L_3^*) + \sigma_{4*}(L_4^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{na} + \rho_{n*}(L) \tag{D-30}$$

$$\geq t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)$$

$$\geq t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L).$$

Then, from (D-23)-(D-26), respectively,

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \tag{D-31}$$

$$\leq t_{2,a-1} + \rho_{2*}(L^*_2) + \sigma_{3*}(L^*_3) + \ldots + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)$$

$$\leq t_{2,a-1} + \rho^*_2(L^*_2) + \sigma^*_3(L^*_3) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \tag{D-32}$$

$$\leq t_{n-2,a-1} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)$$

$$\leq t_{n-2,a-1} + \rho^*_{n-2}(L^*_{n-2}) + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \tag{D-33}$$

$$\leq t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)$$

$$\leq t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L)$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-34)}$$

$$\leq t_{n,a-1} + \rho_{n*}(L)$$

$$\leq t_{n,a-1} + \rho_n^*(L).$$

Further Simplified Validity Conditions

Equations (D-31)-(D-34) may be used to further simplify the Max{•} of (D-21)-(D-26). Similarly, (D-27)-(D-30) may be used to further simplify the Min{•} of (D-21)-(D-26). The results, respectively, are

$$t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-35)}$$

$$\leq t^{0,a-1} \leq t_{na} + \rho_{n*}(L)$$

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{1,a-1} \qquad \text{(D-36)}$$

$$\leq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{2,a-1} \qquad \text{(D-37)}$$

$$\leq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\vdots$$

$$t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \le t^{n-2,a-1} \tag{D-38}$$

$$\le t_{n-2,a-1} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)$$

$$t_{n,a-1} + \rho^*_n(L) \le t^{n-1,a-1} \tag{D-39}$$

$$\le t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L)$$

$$t_{1,a-2} + \rho^*_1(L^*_1) + \sigma^*_2(L^*_2) + \ldots + \sigma^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) \tag{D-40}$$

$$\le t^{n,a-1} \le t_{n,a-1} + \rho_{n*}(L).$$

Realizability Inequalities

It is noted below (2-21) that any control response depending on the measurement $\underline{h}[\underline{x}(L,t^a)]$ must not be required until a time $t \ge t^a + \tau_m$. Since the set points are changed at times $t = t_{ja}$ and the measurements are made at times $t^a = t^{j,a-1}$, it follows that the inequalities

$$\text{Min}\{t_{1a},\ldots,t_{na}\} \ge t^{0,a-1} + \tau_m \tag{D-41}$$

$$\text{Min}\{t_{1a},\ldots,t_{na}\} \ge t^{1,a-1} + \tau_m$$

$$\text{Min}\{t_{1a},\ldots,t_{na}\} \ge t^{2,a-1} + \tau_m$$

$$\vdots$$

$$\text{Min}\{t_{1a},\ldots,t_{na}\} \geq t^{n-2,a-1} + \tau_m$$

$$\text{Min}\{t_{1a},\ldots,t_{na}\} \geq t^{n-1,a-1} + \tau_m$$

$$\text{Min}\{t_{1a},\ldots,t_{na}\} \geq t^{n,a-1} + \tau_m$$

must be satisfied. Now from (4-17),

$$t^{n,a-1} = t^{0,a-2}. \tag{D-42}$$

Thus, $t^{n,a-1}$ must satisfy

$$\text{Min}\{t_{1,a-1},\ldots,t_{n,a-1}\} \geq t^{n,a-1} + \tau_m \tag{D-43}$$

in addition to the condition given in (D-41).

Incorporation of Realizability Inequalities

Now incorporate (D-41)-(D-43) into (D-35)-(D-40). The resulting conditions, respectively, are

$$t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \tag{D-44}$$

$$\leq t^{0,a-1} \leq \text{Min}\{t_{na} + \rho_{n*}(L),\ t_{1a} - \tau_m,\ldots,t_{na} - \tau_m\}$$

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{1,a-1} \qquad \text{(D-45)}$$

$$\leq \text{Min}\{t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\}$$

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{2,a-1} \qquad \text{(D-46)}$$

$$\leq \text{Min}\{t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\}$$

$$\vdots$$

$$t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{n-2,a-1} \qquad \text{(D-47)}$$

$$\leq \text{Min}\{t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\}$$

$$t_{n,a-1} + \rho_n^*(L) \leq t^{n-1,a-1} \qquad \text{(D-48)}$$

$$\leq \text{Min}\{t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L),$$

$$t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\}$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-49)}$$

$$\leq t^{n,a-1} \leq \text{Min}\{t_{n,a-1} + \rho_{n*}(L),\ t_{1,a-1} - \tau_m,\ \ldots,\ t_{n,a-1} - \tau_m\}.$$

Inequalities from Incorporated Conditions

Since τ_m and $\rho_{n*}(L)$ are both non-negative,

$$t_{na} - \tau_m \leq t_{na} + \rho_{n*}(L). \qquad \text{(D-50)}$$

In addition, it follows from (D-44)-(D-47), respectively, that

$$\text{Min}\{t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\} \qquad \text{(D-51)}$$

$$\geq t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\text{Min}\{t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\} \qquad \text{(D-52)}$$

$$\geq t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\text{Min}\{t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\} \tag{D-53}$$

$$\geq t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{3,a-1} + \rho_{3*}(L_3^*) + \sigma_{4*}(L_4^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\text{Min}\{t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\} \tag{D-54}$$

$$\geq t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L).$$

Further Simplified Validity Conditions

Equation (D-50) may be used to simplify (D-44) and (D-49). Similarly, (D-51)-(D-54) may be used to simplify (D-45)-(D-48). The resulting conditions, respectively, are

$$t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \tag{D-55}$$

$$\leq t^{0,a-1} \leq \text{Min}\{t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\}$$

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{1,a-1} \tag{D-56}$$

$$\leq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{2,a-1} \qquad \text{(D-57)}$$

$$\leq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\vdots$$

$$t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \leq t^{n-2,a-1} \qquad \text{(D-58)}$$

$$\leq t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{n,a-1} + \rho_n^*(L) \leq t^{n-1,a-1} \qquad \text{(D-59)}$$

$$\leq t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-60)}$$

$$\leq t^{n,a-1} \leq \text{Min}\{t_{1,a-1} - \tau_m, \ldots, t_{n,a-1} - \tau_m\}.$$

Repetition Rate Inequalities

Consider now the measuring instrument constraint (2-21). To impose this constraint on the above validity conditions, it is necessary to establish the order of the $t^{j,a-1}$. To this end, note from (D-44)-(D-49) that

$$t^{0,a-1} \geq t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-61)}$$

$$\geq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_n^*(L)$$

$$\geq t^{1,a-1}$$

$$t^{1,a-1} \geq t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-62)}$$

$$\geq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\geq t^{2,a-1}$$

$$\vdots$$

$$t^{n-2,a-1} \geq t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L) \qquad \text{(D-63)}$$

$$\geq t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\geq t^{n-1,a-1}$$

$$t^{n-1,a-1} \geq t_{n,a-1} + \rho_n^*(L) \qquad \text{(D-64)}$$

$$\geq t_{n,a-1} + \rho_{n*}(L)$$

$$\geq t^{n,a-1}.$$

It follows from (D-61)-(D-64) that

$$t^{0,a-1} \geq t^{1,a-1} \geq t^{2,a-1} \geq \ldots \geq t^{n-2,a-1} \geq t^{n-1,a-1} \geq t^{n,a-1}. \quad \text{(D-65)}$$

Thus, the requirement (2-21) may be written

$$t^{0,a-1} \geq t^{1,a-1} + \tau_m \quad \text{(D-66)}$$

$$t^{1,a-1} \geq t^{2,a-1} + \tau_m$$

$$t^{2,a-1} \geq t^{3,a-1} + \tau_m$$

$$\vdots$$

$$t^{n-2,a-1} \geq t^{n-1,a-1} + \tau_m$$

$$t^{n-1,a-1} \geq t^{n,a-1} + \tau_m$$

$$t^{n,a-1} = t^{0,a-2} \geq t^{1,a-2} + \tau_m.$$

Incorporation of Repetition Rate Inequalities

Now incorporate the inequalities (D-66) into (D-55)-(D-60). The resulting conditions, respectively, are

$$\text{Max}\{t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \quad \text{(D-67)}$$

$$t^{1,a-1} + \tau_m\} \leq t^{0,a-1}$$

$$\leq \text{Min}\{t_{1a} - \tau_m, \ldots, t_{na} - \tau_m\}$$

$$\text{Max}\{t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \quad \text{(D-68)}$$

$$t^{2,a-1} + \tau_m\} \leq t^{1,a-1}$$

$$\leq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\text{Max}\{t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \quad \text{(D-69)}$$

$$t^{3,a-1} + \tau_m\} \leq t^{2,a-1}$$

$$\leq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\vdots$$

$$\text{Max}\{t_{n-1,a-1} + \rho_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \quad \text{(D-70)}$$

$$t^{n-1,a-1} + \tau_m\} \leq t^{n-2,a-1}$$

$$\leq t_{n-2,a-1} + \rho_{n-2*}(L_{n-2}^*) + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\text{Max}\{t_{n,a-1} + \rho_n^*(L), \quad \text{(D-71)}$$

$$t^{n,a-1} + \tau_m\} \leq t^{n-1,a-1}$$

$$\leq t_{n-1,a-1} + \rho_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L)$$

$$\text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \qquad \text{(D-72)}$$

$$t^{1,a-2} + \tau_m\} \leq t^{n,a-1}$$

$$\leq \text{Min}\{t_{1,a-1} - \tau_m, \ldots, t_{n,a-1} - \tau_m\}.$$

Stage n Sequencing Equations

Consider now the derivation of the set point command time $t_{n,a-1}$ and the measurement sample time $t^{n,a-1}$ from the inequalities (D-67)-(D-72). First note that $t_{n,a-1}$ appears on the right-hand side of only one of these inequalities, namely (D-72). It follows from this inequality that

$$t_{n,a-1} \geq \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) \qquad \text{(D-73)}$$

$$+ \sigma_n^*(L) + \tau_m, t^{1,a-2} + t\,\tau_m\}.$$

To avoid unnecessary delay in system optimization, choose

$$t_{n,a-1} = \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) \qquad \text{(D-74)}$$

$$+ \sigma_n^*(L) + \tau_m, t^{1,a-2} + 2\,\tau_m\}.$$

Now consider $t^{n,a-1}$. Substituting (D-74) into (D-72), the inequality

$$\text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L), \tag{D-75}$$

$$t^{1,a-2} + \tau_m\} \leq t^{n,a-1} \leq$$

$$\text{Min}\{t_{1,a-1} - \tau_m, \ldots, t_{n-1,a-1} - \tau_m,$$

$$\text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L),$$

$$t^{1,a-2} + \tau_m\}\}$$

is obtained. It follows from (D-75) that the only choice for $t^{n,a-1}$ is

$$t^{n,a-1} = \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots \tag{D-76}$$

$$+ \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L),\ t^{1,a-2} + \tau_m\}.$$

Using (D-76), (D-74) can be written

$$t_{n,a-1} = t^{n,a-1} + \tau_m. \tag{D-77}$$

Equations (D-76) and (D-77) are the stage n sequencing equations given in Chapter IV as (4-18) and (4-19).

Stage n-1 Sequencing Equations

Consider now the set point command time $t_{n-1,a-1}$ and the measurement sample time $t^{n-1,a-1}$. The command time $t_{n-1,a-1}$ appears on the right-hand side of only two of the inequalities (D-67)-(D-72), namely (D-71) and (D-72). It follows from these inequalities that

$$t_{n-1,a-1} \geq \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) \quad \text{(D-78)}$$

$$+ \sigma_n^*(L) + \tau_m,\ t^{1,a-2} + 2\tau_m,$$

$$t_{n,a-1} + \rho_n^*(L) - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$t^{n,a-1} + \tau_m - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L)\}$$

$$= \text{Max}\{t_{n,a-1},$$

$$t_{n,a-1} + \rho_n^*(L) - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$t^{n,a-1} + \tau_m - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L)\}.$$

As before, choose

$$t_{n-1,a-1} = \text{Max}\{t_{n,a-1}, \quad \text{(D-79)}$$

$$t_{n,a-1} + \rho_n^*(L) - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$t^{n,a-1} + \tau_m - \rho_{n-1*}(L^*_{n-1}) - \sigma_{n*}(L)\}.$$

Now consider $t^{n-1,a-1}$. Of the range of values permitted by (D-71), choose the one which permits the latest sampling of system states:

$$t^{n-1,a-1} = t_{n-1,a-1} + \rho_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L). \qquad \text{(D-80)}$$

Equations (D-79) and (D-80) are the stage n-1 sequencing equations given in Chapter IV as (4-20) and (4-21).

Stage n-2,...,1 Sequencing Equations

Analyses similar to that given above establish the sequencing equations for the (n-2)nd to the 1st stages. Thus, from (D-72) and (D-70),

$$t_{n-2,a-1} = \text{Max}\{t_{n,a-1}, \qquad \text{(D-81)}$$

$$t_{n-1,a-1} + \rho^*_{n-1}(L^*_{n-1}) + \sigma^*_n(L) - \rho_{n-2*}(L^*_{n-2})$$

$$- \sigma_{n-1*}(L^*_{n-1}) - \sigma_{n*}(L),$$

$$t^{n-1,a-1} + \tau_m - \rho_{n-2*}(L^*_{n-2}) - \sigma_{n-1*}(L^*_{n-1}) - \sigma_{n*}(L)\}$$

and

$$t^{n-2,a-1} = t_{n-2,a-1} + \rho_{n-2*}(L^*_{n-2}) + \sigma_{n-1*}(L^*_{n-1}) + \sigma_{n*}(L). \qquad \text{(D-82)}$$

These are the equations given in Chapter IV as (4-22) and (4-23). Now from (D-72) and (D-69),

$$t_{2,a-1} = \text{Max}\{t_{n,a-1},$$ (D-83)

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$- \rho_{2*}(L_2^*) - \sigma_{3*}(L_3^*) - \ldots - \sigma_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$t^{3,a-1} + \tau_m - \rho_{2*}(L_2^*) - \sigma_{3*}(L_3^*) - \ldots - \sigma_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L)\}$$

and

$$t^{2,a-1} = t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L). \quad \text{(D-84)}$$

These are the equations given in Chapter IV as (4-24) and (4-25). And finally, from (D-72) and (D-68),

$$t_{1,a-1} = \text{Max}\{t_{n,a-1},$$ (D-85)

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) + \sigma_n^*(L)$$

$$- \rho_{1*}(L_1^*) - \sigma_{2*}(L_2^*) - \ldots - \sigma_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$t^{2,a-1} + \tau_m - \rho_{1*}(L_1^*) - \sigma_{2*}(L_2^*) - \ldots - \sigma_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L)\}$$

and

$$t^{1,a-1} = t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{n-1*}(L_{n-1}^*) + \sigma_{n*}(L). \qquad \text{(D-86)}$$

These are the equations given in Chapter IV as (4-26) and (4-27).

APPENDIX E

PROOF OF CONTROL-CONSTRAINED CONVERGENCE

The first step in establishing the convergence of the control-constrained algorithm is to establish the location of the control-constrained optimum. Consider, then, the function to be maximized, as given by (4-8),

$$F(u_{1a},\ldots,u_{na},t^a) = F(u_{1,a-1},\ldots,u_{n,a-1},t^{a-1}) + \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} (u_{ja}-u_{j,a-1}), \tag{E-1}$$

and the constraints to be satisfied, as given by (2-15):

$$u_{j*} \le u_j(t) \le u_j^* \qquad j=1,\ldots,n. \tag{E-2}$$

The set points which maximize $F(u_{1a},\ldots,u_{na},t^a)$ subject to these constraints are

$$u_{ja}^m = u_{j0} + \beta_j \operatorname{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j \tag{E-3}$$

where β_j is the largest non-negative integer for which

$$u_{j*} \le u_{j0} + \beta_j \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j \le u_j^*. \qquad \text{(E-4)}$$

To prove that the control-constrained algorithm drives the u_{ja} to the u_{ja}^m, it is necessary to assume that the signs of the partial derivatives in (E-1) are time-invariant:

$$\text{sign}\left[\left.\frac{\partial F}{\partial u_{j\alpha}}\right|_{\alpha-1}\right] = \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right] \qquad \alpha=2,\ldots,a. \qquad \text{(E-5)}$$

Then, from (4-11), (4-37), and (5-3), the complete u_{ja} sequence is

$$u_{j1} = u_{j0} + 1\,\Delta u_j \qquad \text{(E-6)}$$

$$u_{j2} = u_{j1} + \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j = u_{j0} + \left(1 + \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\right)\Delta u_j$$

$$u_{j3} = u_{j2} + \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j = u_{j0} + \left(1 + 2\,\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\right)\Delta u_j$$

$$\vdots$$

$$u_{j,\alpha_j} = u_{j,\alpha_j-1} + \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j = u_{j0} + \left(1 + (\alpha_j-1)\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\right)\Delta u_j$$

$$u_{j,\alpha_j+1} = u_{j,\alpha_j} - \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j = u_{j0} + \left(1 + (\alpha_j-2)\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\right)\Delta u_j$$

$$u_{j,\alpha_j+2} = u_{j,\alpha_j+1} + \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\Delta u_j = u_{j0} + \left(1 + (\alpha_j - 1)\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\right)\Delta u_j$$

$$\vdots$$

where α_j is the greatest integer for which

$$u_{j*} \le u_{j0} + \left(1 + (\alpha_j - 1)\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]\right)\Delta u_j \le u_j^*. \qquad \text{(E-7)}$$

Thus, the set point u_{ja}^m is achieved after α_j algorithmic cycles where, from (E-4) and (E-7),

$$1 + (\alpha_j - 1)\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right] = \beta_j \text{ sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]. \qquad \text{(E-8)}$$

Then from (E-8),

$$\alpha_j = \frac{\beta_j \text{ sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right] + \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right] - 1}{\text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right]} \qquad \text{(E-9)}$$

$$= \beta_j + 1 - \text{sign}\left[\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\right].$$

And from (E-6), each u_{ja} establishes a limit cycle about its respective u_{ja}^m after α_j algorithmic cycles. This limit cycle has an amplitude Δu_j

and a period 2 T_c where T_c is the algorithmic cycle time

$$T_c = t_{1,a-1} - t_{1,a-2}. \tag{E-10}$$

And finally, it follows that all n of the u_{ja} establish limit cycles about their respective u_{ja}^m after A algorithmic cycles where

$$A = \text{Max}\{\alpha_1,\ldots,\alpha_n\}. \tag{E-11}$$

APPENDIX F

DERIVATION OF CONSTRAINED SEQUENCING EQUATIONS

Many of the results of Appendix D may be used in deriving the constrained sequencing equations, which establish the set point command times $t_{j,a-1}$, the exit measurement sample times $t^{j,a-1}$, and the constraint measurement times $t^{(jk)b,a-1}$.

Constraint Measurements Required

Consider first the measurements required by all j of the approximations given in (5-11):

$$Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)0,a-1}) \tag{F-1}$$

$$Z_{jkr}(u_{1,a-2},u_{2,a-1},\ldots,u_{j,a-1},t^{(jk)1,a-1})$$

$$Z_{jkr}(u_{1,a-2},u_{2,a-2},u_{3,a-1},\ldots,u_{j,a-1},t^{(jk)2,a-1})$$

$$\vdots$$

$$Z_{jkr}(u_{1,a-2},\ldots,u_{j-2,a-2},u_{j-1,a-1},u_{j,a-1},t^{(jk)j-2,a-1})$$

$$Z_{jkr}(u_{1,a-2},\ldots,u_{j-1,a-2},u_{j,a-1},t^{(jk)j-1,a-1})$$

$$Z_{jkr}(u_{1,a-2},\ldots,u_{j,a-2},t^{(jk)j,a-1}).$$

Constraint Measurement Validity Conditions

Associated with each of these measurements is a validity condition which may be obtained from (3-51) by setting ℓ to ℓ_{jk}, t to the required $t^{(jk)b,a-1}$, and "a" to the index corresponding to the set points in the required measurement. These conditions are identical to (D-1)-(D-6) except for the replacement of n by j, L by ℓ_{jk}, and $t^{j,a-1}$ by $t^{(jk)b,a-1}$. It follows that (D-35)-(D-40), which are derived from (D-1)-(D-6), may be used in the present analysis if the same notational changes are made.

A similar argument applies to using (D-55)-(D-60) if τ_m is set to zero, except that the collection of set points $\{u_{1a},\ldots,u_{na}\}$ arising from the realizability conditions is retained intact, the upper member of that set being u_{na} and not u_{ja}.

With these notational changes, the following inequalities may be obtained from (D-55)-(D-60):

$$t_{1,a-1} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk}) \qquad \text{(F-2)}$$

$$\leq t^{(jk)0,a-1} \leq \text{Min}\{t_{1a},\ldots,t_{na}\}$$

$$t_{2,a-1} + \rho_2^*(L_2^*) + \sigma_3^*(L_3^*) + \ldots + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk}) \qquad \text{(F-3)}$$

$$\leq t^{(jk)1,a-1}$$

$$\leq t_{1,a-1} + \rho_{1*}(L_1^*) + \sigma_{2*}(L_2^*) + \ldots + \sigma_{j-1*}(L_{j-1}^*) + \sigma_{j*}(\ell_{jk})$$

$$t_{3,a-1} + \rho_3^*(L_3^*) + \sigma_4^*(L_4^*) + \ldots + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk}) \qquad \text{(F-4)}$$

$$\leq t^{(jk)2,a-1}$$

$$\leq t_{2,a-1} + \rho_{2*}(L_2^*) + \sigma_{3*}(L_3^*) + \ldots + \sigma_{j-1*}(L_{j-1}^*) + \sigma_{j*}(\ell_{jk})$$

$$\vdots$$

$$t_{j-1,a-1} + \rho_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk}) \qquad \text{(F-5)}$$

$$\leq t^{(jk)j-2,a-1}$$

$$\leq t_{j-2,a-1} + \rho_{j-2*}(L_{j-2}^*) + \sigma_{j-1*}(L_{j-1}^*) + \sigma_{j*}(\ell_{jk})$$

$$t_{j,a-1} + \rho_j^*(\ell_{jk}) \qquad \text{(F-6)}$$

$$\leq t^{(jk)j-1,a-1}$$

$$\leq t_{j-1,a-1} + \rho_{j-1*}(L_{j-1}^*) + \sigma_{j*}(\ell_{jk})$$

$$t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk}) \qquad \text{(F-7)}$$

$$\leq t^{(jk)j,a-1} \leq \text{Min}\{t_{1,a-1},\ldots,t_{n,a-1}\}.$$

Establishing the $t_{j,a-1}$ and $t^{j,a-1}$

Consider now the derivation of the set point command times $t_{j,a-1}$ and the exit measurement sample times $t^{j,a-1}$ from the inequalities (D-67)-(D-72) and (F-2)-(F-7). The command time $t_{n,a-1}$ appears on the right-hand side of two of these inequalities, namely (D-72) and (F-7). It follows from these inequalities that

$$t_{n,a-1} \geq \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) \quad \text{(F-8)}$$

$$+ \sigma_n^*(L) + \tau_m,\ t^{1,a-2} + 2\tau_m,$$

$$\max_{j=1}^{n} \max_{k=1}^{p_j} \{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots$$

$$+ \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk})\}\}.$$

When simplified, this inequality is identical to (D-73). Thus, (D-74)-(D-77) are still valid in the state-constrained case and $t^{n,a-1}$ and $t_{n,a-1}$ are still given by (4-18) and (4-19).

Now consider the command time $t_{n-1,a-1}$. It follows from (D-72), (D-71), (F-7), and (F-6) with j=n that

$$t_{n-1,a-1} \geq \text{Max}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots + \sigma_{n-1}^*(L_{n-1}^*) \quad \text{(F-9)}$$

$$+ \sigma_n^*(L) + \tau_m,\ t^{1,a-2} + 2\tau_m,$$

$$t_{n,a-1} + \rho_n^*(L) - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$t^{n,a-1} + \tau_m - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(L),$$

$$\underset{j=1}{\overset{n}{\text{Max}}} \underset{k=1}{\overset{P_j}{\text{Max}}}\{t_{1,a-2} + \rho_1^*(L_1^*) + \sigma_2^*(L_2^*) + \ldots$$

$$+ \sigma_{j-1}^*(L_{j-1}^*) + \sigma_j^*(\ell_{jk})\},$$

$$\underset{k=1}{\overset{P_n}{\text{Max}}}\{t_{n,a-1} + \rho_n^*(\ell_{nk}) - \rho_{n-1*}(L_{n-1}^*) - \sigma_{n*}(\ell_{nk})\}\}.$$

This inequality, as above, may be simplified to (D-78). It follows that (D-79) and (D-80) are still valid and that $t_{n-1,a-1}$ and $t^{n-1,a-1}$ are still given by (4-20) and (4-21).

Analyses similar to that given above establish the continued validity of (4-22)-(4-27). Thus, the command times $t_{j,a-1}$ and sample times $t^{j,a-1}$ are unchanged by the addition of state constraints.

Establishing the $t^{(jk)b,a-1}$

Consider now the constraint measurement times $t^{(jk)b,a-1}$. Of the range of values permitted by (F-2)-(F-7), choose the value which permits the latest measurement of system states. Thus, from (F-7),

$$t^{(jk)j,a-1} = \text{Min}\{t_{1,a-1},\ldots,t_{n,a-1}\}. \qquad \text{(F-10)}$$

An examination of (4-18)-(4-27) reveals that the earliest of the $t_{j,a-1}$ is $t_{n,a-1}$. Thus, (F-10) becomes

$$t^{(jk)j,a-1} = t_{n,a-1}. \tag{F-11}$$

Now from (F-6)-(F-3), respectively,

$$t^{(jk)j-1,a-1} = t_{j-1,a-1} + \rho_{j-1*}(L^{*}_{j-1}) + \sigma_{j*}(\ell_{jk}) \tag{F-12}$$

$$t^{(jk)j-2,a-1} = t_{j-2,a-1} + \rho_{j-2*}(L^{*}_{j-2}) + \sigma_{j-1*}(L^{*}_{j-1}) + \sigma_{j*}(\ell_{jk}) \tag{F-13}$$

$$\vdots$$

$$t^{(jk)2,a-1} = t_{2,a-1} + \rho_{2*}(L^{*}_{2}) + \sigma_{3*}(L^{*}_{3}) + \ldots + \sigma_{j-1*}(L^{*}_{j-1}) \tag{F-14}$$

$$+ \sigma_{j*}(\ell_{jk})$$

$$t^{(jk)1,a-1} = t_{1,a-1} + \rho_{1*}(L^{*}_{1}) + \sigma_{2*}(L^{*}_{2}) + \ldots + \sigma_{j-1*}(L^{*}_{j-1}) \tag{F-15}$$

$$+ \sigma_{j*}(\ell_{jk}).$$

And from (F-1),

$$t^{(jk)0,a-1} = t^{(jk)ja}. \tag{F-16}$$

Equations (F-11)-(F-16) are the constraint measurement times given in Chapter V as (5-12).

APPENDIX G

GENERATION OF PROJECTION MATRIX

Presented in this appendix are procedures for generating the projection matrix M_q.

Definition of the y_i and b_i

Consider first some definitions which permit the constraints of (5-5) to be written in the form of (5-15). Using (5-7) and (5-10), (5-5) may be written in the form

$$z_{jkr^*} \le Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) + \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} (u_{ba}-u_{b,a-1}) \le z^*_{jkr} \qquad \text{(G-1)}$$

$$j=1,\ldots,n \quad k=1,\ldots,p_j \quad r=1,\ldots,q_{jk}.$$

This set of constraints may be written as the two sets of constraints

$$\sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} u_{ba} - \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} u_{b,a-1} + Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) - z_{jkr^*} \ge 0 \qquad \text{(G-2)}$$

$$j=1,\ldots,n \quad k=1,\ldots,p_j \quad r=1,\ldots,q_{jk}$$

and

$$- \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} u_{ba} + \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} u_{b,a-1} \tag{G-3}$$

$$- Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) + z^{*}_{jkr} \geq 0$$

$$j=1,\ldots,n \quad k=1,\ldots,p_j \quad r=1,\ldots,q_{jk}.$$

Then, letting the vector $\underline{y}_i$ have components y_{ib},

$$\underline{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{in} \end{bmatrix}, \tag{G-4}$$

and writing (5-15) in the form

$$\sum_{b=1}^{n} y_{ib} u_{ba} - b_i \geq 0 \qquad i=1,\ldots,k, \tag{G-5}$$

the constraints (G-2) can be put in the form of (G-5) by defining the y_{ib} and b_i to be

$$y_{ib} = \begin{cases} \left.\dfrac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} & b=1,\ldots,j \\ 0 & b=j+1,\ldots,n \end{cases} \tag{G-6}$$

and

$$b_i = \sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} u_{b,a-1} \qquad \text{(G-7)}$$

$$- Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) + z_{jkr}^{*}.$$

Similarly, the constraints (G-3) can be put in the form of (G-5) by defining the y_{ib} and b_i to be

$$y_{ib} = \begin{cases} -\left.\dfrac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} & b=1,\ldots,j \\ 0 & b=j+1,\ldots,n \end{cases} \qquad \text{(G-8)}$$

and

$$b_i = -\sum_{b=1}^{j} \left.\frac{\partial Z_{jkr}}{\partial u_{ba}}\right|_{a-1} u_{b,a-1} \qquad \text{(G-9)}$$

$$+ Z_{jkr}(u_{1,a-1},\ldots,u_{j,a-1},t^{(jk)a-1}) - z_{jkr}^{*}.$$

The index i, embracing both (G-2) and (G-3), ranges from 1 to k where

$$k = 2 \sum_{j=1}^{n} \sum_{k=1}^{p_j} q_{jk}. \qquad \text{(G-10)}$$

The definitions (G-6)-(G-10) permit the constraints of (5-5) to be written in the form of (5-15).

Basic Projection Matrix Equations

Now consider a procedure for generating a projection matrix M_q from a set of linearly independent normalized constraint vectors

$$\underline{y}_i = \frac{\underline{Y}_i}{|\underline{Y}_i|} = \frac{\underline{Y}_i}{\sqrt{\sum_{j=1}^{n} Y_{ij}^2}} . \qquad \text{(G-11)}$$

Define the n × q matrix N_q by

$$N_q = [\underline{y}_1 \cdots \underline{y}_q] \qquad \text{(G-12)}$$

and the projection matrix M_q by[43]

$$M_q = I - N_q(N_q^T N_q)^{-1} N_q^T \qquad \text{(G-13)}$$

where I is the n × n identity matrix. The matrix M_q projects any vector $\underline{u}$ into the intersection of all q constraints:[43]

$$\underline{y}_i^T(M_q\underline{u}) = 0 \qquad i=1,\ldots,q. \qquad \text{(G-14)}$$

The matrix inversion required by (G-13) can be avoided by generating M_q by the recursive relation[43]

$$M_i = M_{i-1} - \frac{(M_{i-1}\underline{y}_i)(M_{i-1}\underline{y}_i)^T}{|M_{i-1}\underline{y}_i|^2} \qquad i=1,\ldots,q \qquad \text{(G-15)}$$

where

$$M_0 = I. \tag{G-16}$$

This relation provides a means of selecting a set of q linearly independent vectors from an arbitrary set of q vectors, for if some $\underline{y}_i$ satisfies

$$M_{i-1}\underline{y}_i = 0 \qquad i=1,\ldots,q \tag{G-17}$$

then $\underline{y}_i$ is linearly dependent on the vectors $\underline{y}_1,\ldots,\underline{y}_{i-1}$ and can be deleted from the set of q vectors.[43] If M_q is generated in this manner, M_q will project any vector $\underline{u}$ into the intersection of all q constraints:[43]

$$\underline{y}_i^T(M_q\underline{u}) = 0 \qquad i=1,\ldots,q. \tag{G-18}$$

In developing M_q, the $\underline{y}_i$ next entered in the recursive relation (G-15) is that $\underline{y}_i$ for which the quantity θ_i'' is negative and, further, is more negative than the θ_i'' of any other $\underline{y}_i$, where[43]

$$\theta_i'' = \underline{\phi}_{i-1}''^T\underline{y}_i \tag{G-19}$$

and

$$\underline{\phi}_{i-1}'' = \frac{M_{i-1}\underline{g}}{\left|M_{i-1}\underline{g}\right|}. \tag{G-20}$$

This procedure insures that only those constraints which are oriented so as to preclude movement in the gradient direction, or in the evolving projected gradient direction, are included in the projection matrix.

Modified Projection Matrix Equations

The above procedures may be modified to eliminate division and square root operations. Consider, in this regard, the recursive relation

$$M_i = |M_{i-1}\underline{y}_i|^2 M_{i-1} - (M_{i-1}\underline{y}_i)(M_{i-1}\underline{y}_i)^T \prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 \quad i=1,\ldots,q \qquad \text{(G-21)}$$

where

$$M_0 = I. \qquad \text{(G-22)}$$

This relation can be shown to establish a projection matrix M_q which is a scalar multiple of the projection matrix M_q:

$$M_q = \prod_{k=1}^{q} |M_{k-1}\underline{y}_k|^2 M_q. \qquad \text{(G-23)}$$

Specifically, since

$$M_0 = M_0, \qquad \text{(G-24)}$$

(G-23) may be established by showing that

$$M_i = \prod_{k=1}^{i} |M_{k-1}\underline{y}_k|^2 M_i \qquad \text{(G-25)}$$

given that

$$M_{i-1} = \prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 M_{i-1}. \tag{G-26}$$

Equation (G-25) may be easily established by inserting (G-11), (G-15), and (G-26) into (G-21):

$$\begin{aligned} M_i &= |M_{i-1}\underline{y}_i|^2 \prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 M_{i-1} \\ &\quad - (M_{i-1}\underline{y}_i)(M_{i-1}\underline{y}_i)^T \prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 \\ &= |M_{i-1}\underline{y}_i|^2 \prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 \left(M_{i-1} - \frac{(M_{i-1}\underline{y}_i)(M_{i-1}\underline{y}_i)^T}{|M_{i-1}\underline{y}_i|^2} \right) \end{aligned} \tag{G-27}$$

$$= \prod_{k=1}^{i} |M_{k-1}\underline{y}_k|^2 \times$$

$$\left(M_{i-1} - \frac{\left(\prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 M_{i-1} |\underline{y}_i| \underline{u}_i \right) \left(\prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 M_{i-1} |\underline{y}_i| \underline{u}_i \right)^T}{\left| \prod_{k=1}^{i-1} |M_{k-1}\underline{y}_k|^2 M_{i-1} |\underline{y}_i| \underline{u}_i \right|^2} \right)$$

$$= \prod_{k=1}^{i} |M_{k-1}\underline{y}_k|^2 \left(M_{i-1} - \frac{(M_{i-1}\underline{u}_i)(M_{i-1}\underline{u}_i)^T}{|M_{i-1}\underline{u}_i|^2} \right)$$

$$= \prod_{k=1}^{i} |M_{k-1}\underline{y}_k|^2 M_i.$$

Now consider the problem of determining which of several θ_i'' is the most negative. Define the unnormalized functions

$$\theta_i = \underline{\phi}_{i-1}^T \underline{y}_i \tag{G-28}$$

and

$$\underline{\phi}_{i-1} = M_{i-1}\underline{g}. \tag{G-29}$$

It follows from (G-11), (G-19)-(G-20), and (G-25) that

$$\theta_i'' = \frac{\underline{\phi}_{i-1}^T}{|\underline{\phi}_{i-1}|} \frac{\underline{y}_i}{|\underline{y}_i|} = \frac{\theta_i}{|\underline{\phi}_{i-1}||\underline{y}_i|}. \tag{G-30}$$

From (G-30), the sign of θ_i'' is the same as the sign of θ_i. Of the negative θ_i, the one corresponding to the most negative θ_i'' is the one with the largest $(\theta_i'')^2$ where

$$(\theta_i'')^2 = \frac{(\theta_i)^2}{\sum_{j=1}^{n} (\phi_{i-1,j})^2 \sum_{j=1}^{n} (y_{ij})^2}. \tag{G-31}$$

More simply, the one with the largest $(\theta_i'')^2$ is the one with the largest

$$\frac{(\theta_i)^2}{\sum_{j=1}^{n} (y_{ij})^2}. \tag{G-32}$$

Finally, the divisions required by (G-32) as a prerequisite to deciding whether $(\theta_i'')^2$ or $(\theta_i'')^2$ is larger may be avoided by replacing the inequality

$$\frac{(\theta_i)^2}{\sum_{j=1}^{n} (y_{ij})^2} > \frac{(\theta_i)^2}{\sum_{j=1}^{n} (y_{ij})^2} \tag{G-33}$$

with the inequality

$$(\theta_i)^2 \sum_{j=1}^{n} (y_{ij})^2 > (\theta_i)^2 \sum_{j=1}^{n} (y_{ij})^2. \tag{G-34}$$

Thus, division and square root operations may be completely avoided in the generation of the projection matrix M_q.

APPENDIX H

PROOF OF STATE-CONSTRAINED CONVERGENCE

The proof of the convergence of the state-constrained algorithm is divided into several sections. First, the d_k of (5-24) are established. Next, the β^*-tuple $\underline{\pi}$ is defined. Then, the minimization problem which defines $\overset{''}{\beta}$ is extended, and the uniqueness of $\overset{''}{\beta}$ is established for the case q=n-1. Then, for unique $\overset{''}{\beta}$, the convergence of the state-constrained algorithm is established. Finally, for non-unique $\overset{''}{\beta}$, the basic algorithm is extended and the convergence of the extended procedure established.

Proof of Equation (5-24)

The d_k given by (5-24) may be established by showing that each component of the linear combination (5-23), after insertion of (5-22) and (5-24), is equal to w_j. Consider then the jth component of (5-23) when $j=\psi_1$:

$$d_1 c_{1j} + d_2 c_{2j} + \ldots + d_n c_{nj} \tag{H-1}$$

$$= 0.5\ \mathrm{SGN}\left[\sum_{k=2}^{n} \frac{|w_{\psi_k}|}{\Delta u_{\psi_k}} - (n-3)\,\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}}\right] \mathrm{SGN}\ \mathrm{sign}(w_j)\Delta u_j$$

$$+ 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_2}|}{\Delta u_{\psi_2}}\right] \mathrm{sign}(w_j)\Delta u_j$$

$$\vdots$$

$$+ 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_n}|}{\Delta u_{\psi_n}}\right]\text{sign}(w_j)\Delta u_j$$

$$= 0.5[(n-1)-(n-3)]\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}}\text{ sign}(w_j)\Delta u_j$$

$$= |w_j|\text{sign}(w_j)$$

$$= w_j.$$

Consider also the jth component of (5-23) when $j=\psi_b$, b=2,...,n:

$$d_1c_{1j} + d_2c_{2j} + \ldots + d_nc_{nj} \qquad \text{(H-2)}$$

$$= 0.5\ \text{SGN}\left[\sum_{k=2}^{n}\frac{|w_{\psi_k}|}{\Delta u_{\psi_k}} - (n-3)\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}}\right]\text{SGN sign}(w_j)\Delta u_j$$

$$+ 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_2}|}{\Delta u_{\psi_2}}\right]\text{sign}(w_j)\Delta u_j$$

$$\vdots$$

$$+ 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_{b-1}}|}{\Delta u_{\psi_{b-1}}}\right]\text{sign}(w_j)\Delta u_j$$

$$- 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_b}|}{\Delta u_{\psi_b}}\right]\text{sign}(w_j)\Delta u_j$$

$$+ 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_{b+1}}|}{\Delta u_{\psi_{b+1}}}\right]\text{sign}(w_j)\Delta u_j$$

$$\vdots$$

$$+ 0.5\left[\frac{|w_{\psi_1}|}{\Delta u_{\psi_1}} - \frac{|w_{\psi_n}|}{\Delta u_{\psi_n}}\right]\text{sign}(w_j)\Delta u_j$$

$$= 0.5\left[\frac{|w_{\psi_b}|}{\Delta u_{\psi_b}} + \frac{|w_{\psi_b}|}{\Delta u_{\psi_b}}\right]\text{sign}(w_j)\Delta u_j$$

$$= |w_j|\text{sign}(w_j)$$

$$= w_j.$$

Since each of the above is equal to w_j, (5-24) is established.

Definition of the β^*-tuple $\underline{\pi}$

Consider now the ordered β^*-tuple $\underline{\pi} = (\pi_0,\ldots,\pi_{\beta^*-1})$ which is used to arrange the $\underline{c}_k$ in the tracking cycle. For any particular k, the

number of $\underline{c}_k$ is Kd_k. These Kd_k vectors may be uniformly distributed throughout the tracking cycle by arranging the β^* non-negative quantities

$$\frac{(\lambda_k-1)}{Kd_k} \qquad k=1,\ldots,n \qquad \lambda_k=1,\ldots,Kd_k \tag{H-3}$$

in ascending order, arranging by ascending index "k" in case of ties, and defining the ordered β^*-tuple

$$\underline{\pi} = (\pi_0,\ldots,\pi_{\beta^*-1}) \tag{H-4}$$

to be the ordered list of "k" indices of the arranged $(\lambda_k-1)/Kd_k$. Thus, π_0 is the "k" index of the smallest $(\lambda_k-1)/Kd_k$ and π_{β^*-1} is the "k" index of the largest.

If K is the smallest positive constant for which all of the products $Kd_1,\ldots,Kd_n$ are integers, the smallest $\lambda_k>1$ for which there can be an n-way tie is $\lambda_k = Kd_k + 1$, which is not in the range of λ_k. But if K is larger than required by a factor of K^*, the smallest $\lambda_k>1$ for which there can be an n-way tie is $\lambda_k = (Kd_k/K^*) + 1$, which *is* in the range of λ_k. If such a tie occurs, K should be replaced by K/K^*.

Extended Procedure for Selecting β''

In case the minimization problem (5-34)-(5-35) does not have a unique solution, let β'' be that β' which minimizes the sum

$$\sum_{i=1}^{q-1} E_{\beta_i} \qquad \beta' = \begin{cases} \text{solutions to original} \\ \text{minimization problem} \end{cases} \tag{H-5}$$

subject to the constraints (5-35). Any multiple solutions of this minimization problem must have equal E_{β_q}. Continuing in this manner until β'' is that β' which minimizes the sum

$$\sum_{i=1}^{1} E_{\beta_i} \qquad \beta' = \begin{cases} \text{solutions to above} \\ \text{minimization problem} \end{cases} \tag{H-6}$$

subject to the constraints (5-35), any multiple solutions of this minimization problem must have equal E_{β_2} and, furthermore, equal E_{β_1}. Because of the requirement of equal E_{β_i}, the above sequence of minimization problems usually terminates in a unique β''.

Proof of Unique β'' When q=n-1

When q=n-1, the above is guaranteed to terminate in a unique β''. To establish this uniqueness, suppose that β''^1 and β''^2 are both solutions, where $\beta''^2 > \beta''^1$. To simplify notation, let B_{β_i} be that β which is the solution to the minimization problem (5-33). Using this notation, (5-33) may be written

$$E_{\beta_i} = \underline{y}_i^T \underline{u}_{\underline{a}+B_{\beta_i},\beta'} - b_i. \tag{H-7}$$

Since

$$E_{\beta''^1 i} = E_{\beta''^2 i} \qquad i=1,\ldots,q, \tag{H-8}$$

it follows from (H-7) that

$$\underline{y}_i^T \underline{u}_{\underline{a}+B_{\beta''^1 i},\beta''^1} = \underline{y}_i^T \underline{u}_{\underline{a}+B_{\beta''^2 i},\beta''^2}. \tag{H-9}$$

Inserting (5-32) into (H-9),

$$\underline{y}_i^T \sum_{b=0}^{B_{\beta''^1 i}} \underline{c}_{\pi_{b+\beta''^1}} = \underline{y}_i^T \sum_{b=0}^{B_{\beta''^2 i}} \underline{c}_{\pi_{b+\beta''^2}}. \tag{H-10}$$

Then with summation indices rearranged,

$$\underline{y}_i^T \sum_{b=\beta''^1}^{\beta''^2-1} \underline{c}_{\pi_b} + \underline{y}_i^T \sum_{b=\beta''^2}^{B_{\beta''^1 i}+\beta''^1} \underline{c}_{\pi_b} \tag{H-11}$$

$$= \underline{y}_i^T \sum_{b=\beta''^2}^{B_{\beta''^2 i}+\beta''^2} \underline{c}_{\pi_b}.$$

That point in the tracking cycle which sweeps closest to the ith constraint boundary is the same no matter what the alignment. Thus, from (5-32),

$$b + \overset{'}{\beta}\Big|_{\substack{b = B_{\overset{''}{\beta}^1 i} \\ \overset{'}{\beta} = \overset{''}{\beta}^1}} = b + \overset{'}{\beta}\Big|_{\substack{b = B_{\overset{''}{\beta}^2 i} \\ \overset{'}{\beta} = \overset{''}{\beta}^2}} . \tag{H-12}$$

That is,

$$B_{\overset{''}{\beta}^1 i} + \overset{''}{\beta}^1 = B_{\overset{''}{\beta}^2 i} + \overset{''}{\beta}^2. \tag{H-13}$$

Then from (H-11),

$$\underline{y}_i^T \sum_{b=\overset{''}{\beta}^1}^{\overset{''}{\beta}^2-1} \underline{c}_{\pi_b} = 0, \tag{H-14}$$

which implies that the vector $\underline{u}$ defined by

$$\underline{u} = \sum_{b=\overset{''}{\beta}^1}^{\overset{''}{\beta}^2-1} \underline{c}_{\pi_b} \tag{H-15}$$

is parallel to all q constraint boundaries and, consequently, in the 1-dimensional Euclidean space (1=n-q) defined by their intersection. Since $\underline{w}$ is also in this 1-dimensional space, $\underline{u}$ is parallel to $\underline{w}$. But if K is the smallest positive constant for which the products $Kd_1,\ldots,Kd_n$ are all integers, $\underline{u}$ can be parallel to $\underline{w}$ if and only if

$$\overset{''}{\beta}^2 - 1 = \overset{''}{\beta}^1 + (\beta^* - 1). \tag{H-16}$$

That is, if and only if

$$\beta''^2 - \beta''^1 = \beta^*. \tag{H-17}$$

However, this is not possible, for $\beta'=0,\ldots,\beta^*-1$ is the range from which β''^2 and β''^1 are taken. Thus, there is a unique β'' for q=n-1.

Proof of Convergence for Unique β''

The convergence of the state-constrained algorithm may be established by showing that

$$\underline{u}_{a+\beta^*-1} = \underline{u}_{a-1} + K\underline{w}. \tag{H-18}$$

To establish this relation, it is necessary to assume, in addition to a unique β'', that the gradient $\underline{g}$ and the q constraint boundaries used to generate the projection matrix M_q are time-invariant over the algorithmic cycles $a,\ldots,a+\beta^*-1$. It is also necessary to introduce a more complex notation for $E_{\beta' i}$ and β'', namely $E^a_{\beta' i}$ and β''^a, and to assume that

$$E^a_{\beta''^a{}_i} = 0 \qquad i=1,\ldots q. \tag{H-19}$$

To begin the proof of (H-18), use (5-32) to obtain

$$\underline{u}_{a+B_{\beta''^a{}_i},\beta''^a} = \underline{u}_{a-1} + \sum_{b=0}^{B_{\beta''^a{}_i}} c_{\pi_{b+\beta''^a}} \tag{H-20}$$

$$= \underline{u}_{a-1} + \underline{c}_{\pi_{\beta}{}^{"}a} + \sum_{b=1}^{B_{\beta}{}^{"}a_i} \underline{c}_{\pi_{b+\beta}{}^{"}a}.$$

Then rewrite (5-36) in the more complex notation

$$\underline{u}_a = \underline{u}_{a-1} + \underline{c}_{\pi_{\beta}{}^{"}a}. \tag{H-21}$$

Inserting (H-21) into (H-20),

$$\underline{u}_{a+B_{\beta}{}^{"}a_i,\beta}{}^{"}a = \underline{u}_a + \sum_{b=1}^{B_{\beta}{}^{"}a_i} \underline{c}_{\pi_{b+\beta}{}^{"}a} \tag{H-22}$$

$$= \underline{u}_a + \sum_{b=0}^{B_{\beta}{}^{"}a_i - 1} \underline{c}_{\pi_{b+\beta}{}^{"}a+1}.$$

Now from (5-32),

$$\underline{u}_{a+1+B_{\beta}{}^{"}a+1_i,\beta}{}^{"}a+1 = \underline{u}_a + \sum_{b=0}^{B_{\beta}{}^{"}a+1_i} \underline{c}_{\pi_{b+\beta}{}^{"}a+1}. \tag{H-23}$$

But, as in (H-13),

$$B_{\beta}{}^{"}a+1_i + \beta^{"}a+1 = B_{\beta}{}^{"}a_i + \beta^{"}a. \tag{H-24}$$

Thus, (H-23) becomes

$$\underline{u}_{a+1+B_{\beta^{\alpha_{a+1}}},\beta^{\alpha_{a+1}}} = \underline{u}_{a} + \sum_{b=0}^{B_{\beta^{\alpha_i}}+\beta^{\alpha_a}-\beta^{\alpha_{a+1}}} \underline{c}_{\pi_{b+\beta^{\alpha_{a+1}}}}. \tag{H-25}$$

It then follows that the alignment

$$\beta^{\alpha_{a+1}} = \beta^{\alpha_a} + 1 \tag{H-26}$$

is the solution to the minimization problem (5-34)-(5-35) at the (a+1)st algorithmic cycle, for with this alignment all q of the $E'^{a+1}_{\beta i}$ are zero. That is,

$$E^{a+1}_{\beta^{\alpha_{a+1}}} = \underline{y}_i^T \underline{u}_{a+1+B_{\beta^{\alpha_{a+1}}},\beta^{\alpha_{a+1}}} - b_i, \tag{H-27}$$

and if (H-26) is satisfied, a comparison of (H-22) and (H-25) yields

$$\underline{u}_{a+B_{\beta^{\alpha_i}},\beta^{\alpha_a}} = \underline{u}_{a+1+B_{\beta^{\alpha_{a+1}}},\beta^{\alpha_{a+1}}}. \tag{H-28}$$

Then inserting (H-28) into (H-27),

$$E^{a+1}_{\beta^{\alpha_{a+1}}} = \underline{y}_i^T \underline{u}_{a+B_{\beta^{\alpha_i}},\beta^{\alpha_a}} - b_i \tag{H-29}$$

$$= E^{a}_{\pi_{\beta} a_i}$$

$$= 0$$

$$i=1,\ldots,q.$$

Finally, it follows from (H-21) that

$$\underline{u}_{a+\beta^*-1} = \underline{u}_{a+\beta^*-2} + \underline{c}_{\pi_{\beta} a+\beta^*-1} \tag{H-30}$$

$$= \underline{u}_{a+\beta^*-3} + \underline{c}_{\pi_{\beta} a+\beta^*-2} + \underline{c}_{\pi_{\beta} a+\beta^*-1}$$

$$= \cdots$$

$$= \underline{u}_{a-1} + \sum_{b=0}^{\beta^*-1} \underline{c}_{\pi_{\beta} a+b}.$$

Then inserting (H-26) into (H-30),

$$\underline{u}_{a+\beta^*-1} = \underline{u}_{a-1} + \sum_{b=0}^{\beta^*-1} \underline{c}_{\pi_{\beta} a+b} \tag{H-31}$$

$$= \underline{u}_{a-1} + \sum_{b=0}^{\beta^*-1} \underline{c}_{\pi_b}$$

$$= \underline{u}_{a-1} + \sum_{k=1}^{n} Kd_k \underline{c}_k$$

$$= \underline{u}_{a-1} + K\underline{w}.$$

Thus, (H-18) is established.

Modifications for Non-Unique $\overset{''}{\beta}$

The uniqueness proof given by (H-7)-(H-17) breaks down when $q<n-1$, for the vectors $\underline{u}$ and $\underline{w}$, while in the same $(n-q)$-dimensional space, are not necessarily parallel. Thus, when $q<n-1$, a non-unique $\overset{''}{\beta}$ is possible, though not likely.

For completeness, however, assume that $\overset{''2}{\beta} > \overset{''1}{\beta}$ are both solutions to the minimization problem terminating in (H-6). Separate the β^*-tuple $\underline{\pi}$ into $\underline{\pi}^1$ and $\underline{\pi}^2$,

$$\underline{\pi}^1 = (\pi_0,\ldots,\pi_{\overset{''1}{\beta}},\pi_{\overset{''2}{\beta}+1},\ldots,\pi_{\beta^*-1}) \qquad \text{(H-32)}$$

$$\underline{\pi}^2 = (\pi_{\overset{''1}{\beta}+1},\ldots,\pi_{\overset{''2}{\beta}}),$$

and consider the two vectors $K\underline{w}^1$ and $K\underline{w}^2$:

$$K\underline{w}^1 = \sum_{\substack{\beta=0 \\ \pi_\beta \in \underline{\pi}^1}}^{\beta^*-1} \underline{c}_{\pi_\beta} \qquad \text{(H-33)}$$

$$K\underline{w}^2 = \sum_{\substack{\beta=0 \\ \pi_\beta \in \underline{\pi}^2}}^{\beta^*-1} \underline{c}_{\pi_\beta}.$$

One or both of these vectors has a positive component in the direction of the gradient,

$$(K\underline{w}^1)^T\underline{g} > 0 \quad \text{and/or} \quad (K\underline{w}^2)^T\underline{g} > 0, \tag{H-34}$$

since their sum

$$K\underline{w} = K\underline{w}^1 + K\underline{w}^2 \tag{H-35}$$

satisfies[43]

$$(K\underline{w})^T\underline{g} > 0. \tag{H-36}$$

Thus, with n defined by

$$n = \begin{cases} 1 & (K\underline{w}^1)^T\underline{g} \geq (K\underline{w}^2)^T\underline{g} \\ 2 & (K\underline{w}^1)^T\underline{g} < (K\underline{w}^2)^T\underline{g} \end{cases} \tag{H-37}$$

and the integers λ_k^n defined as the number of elements of $\underline{\pi}^n$ equal to "k," let β^{*n} be the total number of elements in $\underline{\pi}^n$:

$$\beta^{*n} = \sum_{k=1}^{n} \lambda_k^n. \tag{H-38}$$

Then arrange the β^{*n} non-negative quantities

$$\frac{(\lambda_k-1)}{\lambda_k^n} \qquad k=1,\ldots,n \qquad \lambda_k=1,\ldots,\lambda_k^n \tag{H-39}$$

in ascending order, and proceed as before. This modified procedure will track the vector $K\underline{w}^n$ in β^{*n} algorithmic cycles rather than the vector $K\underline{w}$ in β^* algorithmic cycles. Convergence to the optimum, or to a set point for which q=n-1, is assured by (H-34).

APPENDIX I

DERIVATION OF CONSTRAINT SET EQUATIONS

Given below are analyses establishing the constraint set equations (5-37)-(5-39). First, an upper bound on the width of the tracking cycle is established, and then, using this bound, (5-38) and (5-39) are shown to satisfy the conditions given below (5-37).

Upper Bound on Tracking Cycle Width

Let γ_i be the width of the tracking cycle $\underline{u}_{a+\beta,\beta}$, (see 5-32) as measured along $\underline{y}_i$:

$$\gamma_i = \underset{\beta=0}{\overset{\beta^*-1}{\text{Max}}} \{\underline{y}_i^T \underline{u}_{a+\beta,\beta}\} - \underset{\beta=0}{\overset{\beta^*-1}{\text{Min}}} \{\underline{y}_i^T \underline{u}_{a+\beta,\beta}\}. \tag{I-1}$$

Since the alignment β has no effect on γ_i, it may be set to zero. Then from (5-32),

$$\gamma_i = \underset{\beta=0}{\overset{\beta^*-1}{\text{Max}}} \{\underline{y}_i^T \sum_{b=0}^{\beta} \underline{c}_{\pi_b}\} - \underset{\beta=0}{\overset{\beta^*-1}{\text{Min}}} \{\underline{y}_i^T \sum_{b=0}^{\beta} \underline{c}_{\pi_b}\}. \tag{I-2}$$

Now from (5-25),

$$\sum_{k=1}^{n} Kd_k \underline{c}_k = K\underline{w}, \tag{I-3}$$

and from (G-14),

$$\underline{y}_i^T(K\underline{w}) = 0. \tag{I-4}$$

Combining (I-3) and (I-4),

$$\underline{y}_i^T \sum_{k=1}^{n} Kd_k \underline{c}_k = 0. \tag{I-5}$$

Then multiplying (I-5) by β/β^*,

$$\underline{y}_i^T \sum_{k=1}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k = 0. \tag{I-6}$$

And combining (I-2) and (I-6),

$$\gamma_i = \underset{\beta=0}{\overset{\beta^*-1}{\text{Max}}} \left\{ \underline{y}_i^T \left[\sum_{b=0}^{\beta} \underline{c}_{\pi_b} - \sum_{k=1}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k \right] \right\} \tag{I-7}$$

$$- \underset{\beta=0}{\overset{\beta^*-1}{\text{Min}}} \left\{ \underline{y}_i^T \left[\sum_{b=0}^{\beta} \underline{c}_{\pi_b} - \sum_{k=1}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k \right] \right\}$$

$$= \underset{\beta=0}{\overset{\beta^*-1}{\text{Max}}} \left\{ \underline{y}_i^T \left[\sum_{\substack{b=0 \\ b\epsilon B_p}}^{\beta} \underline{c}_{\pi_b} + \sum_{\substack{b=0 \\ b\epsilon B_n}}^{\beta} \underline{c}_{\pi_b} - \sum_{\substack{k=1 \\ k\epsilon K_p}}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k - \sum_{\substack{k=1 \\ k\epsilon K_n}}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k \right] \right\}$$

$$- \underset{\beta=0}{\overset{\beta^*-1}{\text{Min}}} \left\{ \underline{y}_i^T \left[\sum_{\substack{b=0 \\ b\epsilon B_p}}^{\beta} \underline{c}_{\pi_b} + \sum_{\substack{b=0 \\ b\epsilon B_n}}^{\beta} \underline{c}_{\pi_b} - \sum_{\substack{k=1 \\ k\epsilon K_p}}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k - \sum_{\substack{k=1 \\ k\epsilon K_n}}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k \right] \right\}$$

where B_p and B_n are the sets of "b" indices

$$B_p = \{b | \underline{y}_i^T \underline{c}_{\pi_b} \geq 0\}$$
$$B_n = \{b | \underline{y}_i^T \underline{c}_{\pi_b} < 0\} \tag{I-8}$$

and K_p and K_n are the corresponding sets of "k" indices

$$K_p = \{k | \underline{y}_i^T \underline{c}_k \geq 0\}$$
$$K_n = \{k | \underline{y}_i^T \underline{c}_k < 0\}. \tag{I-9}$$

Now consider the first $\beta+1$ elements of the β^*-tuple $\underline{\pi}$ (see H-4), and let A_k be the total number of these elements equal to "k." It follows that

$$\beta + 1 = \sum_{k=1}^{n} A_k. \tag{I-10}$$

If A is the smallest real number for which

$$\frac{(A_k - 1)}{Kd_k} \leq A \qquad k=1,\ldots,n, \tag{I-11}$$

it follows from (H-3) that

$$A_k = \left[\frac{A}{1/Kd_k}\right] + C \tag{I-12}$$

where $[\cdot]$ is the greatest integer function and C is 1 if

$$\frac{A}{1/Kd_k} \neq \left[\frac{A}{1/Kd_k}\right] \tag{I-13}$$

and may be 0 or 1 if

$$\frac{A}{1/Kd_k} = \left[\frac{A}{1/Kd_k}\right]. \tag{I-14}$$

Combining (I-10) and (I-12),

$$\beta + 1 = \sum_{k=1}^{n}\left(\left[\frac{A}{1/Kd_k}\right] + C\right). \tag{I-15}$$

Then using the relation

$$x - 1 < [x] \leq x \tag{I-16}$$

and carefully considering the value of C, it follows from (I-15) that

$$\beta + 1 \leq \sum_{k=1}^{n}\left(\frac{A}{1/Kd_k} + 1\right) = A\beta^* + n \tag{I-17}$$

and

$$\beta + 1 \geq \sum_{k=1}^{n} \left(\frac{A}{1/Kd_k} \right) = A\beta^* . \tag{I-18}$$

Combining (I-17) and (I-18),

$$\frac{\beta + 1 - n}{\beta^*} \leq A \leq \frac{\beta + 1}{\beta^*} . \tag{I-19}$$

Combining (I-12) and (I-19), and again carefully considering the value of C,

$$A_k \leq \frac{A}{1/Kd_k} + 1 \leq \frac{\beta + 1}{\beta^*/Kd_k} + 1 \tag{I-20}$$

and

$$A_k \geq \frac{A}{1/Kd_k} \geq \frac{\beta + 1 - n}{\beta^*/Kd_k} . \tag{I-21}$$

Now inserting (I-20) and (I-21) into (I-7),

$$\gamma_i \leq \max_{\beta=0}^{\beta^*-1} \left\{ \underline{y}_i^T \left[\sum_{\substack{k=1 \\ k\epsilon K_p}}^{n} \left(\frac{\beta + 1}{\beta^*/Kd_k} + 1 \right) \underline{c}_k + \sum_{\substack{k=1 \\ k\epsilon K_n}}^{n} \frac{\beta + 1 - n}{\beta^*/Kd_k} \underline{c}_k \right. \right. \tag{I-22}$$

$$\left. \left. - \sum_{\substack{k=1 \\ k\epsilon K_p}}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k - \sum_{\substack{k=1 \\ k\epsilon K_n}}^{n} \frac{\beta}{\beta^*/Kd_k} \underline{c}_k \right] \right\}$$

$$= -\min_{\beta=0}^{\beta^*-1}\left\{\underline{y}_i^T\left[\sum_{\substack{k=1\\ k\epsilon K_p}}^{n}\frac{\beta+1-n}{\beta^*/Kd_k}\underline{c}_k+\sum_{\substack{k=1\\ k\epsilon K_n}}^{n}\left(\frac{\beta+1}{\beta^*/Kd_k}+1\right)\underline{c}_k\right.\right.$$

$$\left.\left.-\sum_{\substack{k=1\\ k\epsilon K_p}}^{n}\frac{\beta}{\beta^*/Kd_k}\underline{c}_k-\sum_{\substack{k=1\\ k\epsilon K_n}}^{n}\frac{\beta}{\beta^*/Kd_k}\underline{c}_k\right]\right\}$$

$$= \max_{\beta=0}^{\beta^*-1}\left\{\underline{y}_i^T\left[\sum_{\substack{k=1\\ k\epsilon K_p}}^{n}\left(\frac{1}{\beta^*/Kd_k}+1\right)\underline{c}_k+\sum_{\substack{k=1\\ k\epsilon K_n}}^{n}\frac{1-n}{\beta^*/Kd_k}\underline{c}_k\right]\right\}$$

$$= -\min_{\beta=0}^{\beta^*-1}\left\{\underline{y}_i^T\left[\sum_{\substack{k=1\\ k\epsilon K_p}}^{n}\frac{1-n}{\beta^*/Kd_k}\underline{c}_k+\sum_{\substack{k=1\\ k\epsilon K_n}}^{n}\left(\frac{1}{\beta^*/Kd_k}+1\right)\underline{c}_k\right]\right\}$$

$$= \sum_{k=1}^{n}\left(1+\frac{n}{\beta^*/Kd_k}\right)\left|\underline{y}_i^T\underline{c}_k\right|$$

$$= \sum_{k=1}^{n}\left(1+n\,\frac{Kd_k}{\beta^*}\right)\left|\sum_{j=1}^{n}y_{ij}c_{kj}\right|.$$

Now let $\overset{n}{k}$ be that index k for which

$$\frac{Kd_{\overset{\text{"}}{k}}}{\beta^*} \geq \frac{Kd_k}{\beta^*} \qquad k=1,\ldots,\overset{\text{"}}{k}-1,\overset{\text{"}}{k}+1,\ldots,n. \tag{I-23}$$

It follows from (5-26) that

$$\frac{Kd_k}{\beta^*} \leq 1 \qquad k=\overset{\text{"}}{k} \tag{I-24}$$

and

$$\frac{Kd_k}{\beta^*} \leq \frac{1}{2} \qquad k=1,\ldots,\overset{\text{"}}{k}-1,\overset{\text{"}}{k}+1,\ldots,n. \tag{I-25}$$

For suppose one of the latter exceeds 1/2. Then all of the others must be less than 1/2, and that violates the assumption that the $\overset{\text{"}}{k}$th one is the largest. Thus, inserting (I-24) and (I-25) into (I-22),

$$\gamma_i \leq (1+n)\left|\sum_{j=1}^{n} y_{ij}c_{\overset{\text{"}}{k}j}\right| + \sum_{\substack{k=1 \\ k\neq\overset{\text{"}}{k}}}^{n}\left(1+\frac{n}{2}\right)\left|\sum_{j=1}^{n} y_{ij}c_{kj}\right|. \tag{I-26}$$

Now let J be that index j for which

$$|y_{iJ}|\Delta u_J \geq |y_{ij}|\Delta u_j \qquad j=1,\ldots,J-1,J+1,\ldots,n. \tag{I-27}$$

The right-hand side of (I-26) will take on its greatest value if the result of (5-22) is

$$c_{kj} = \begin{cases} \text{sign}(y_{ij})\Delta u_j & k=\overset{"}{k} \qquad\qquad\qquad\qquad j=1,\ldots,n \\ \text{sign}(y_{ij})\Delta u_j & k=1,\ldots,\overset{"}{k}-1,\overset{"}{k}+1,\ldots,n \quad j=1,\ldots,\overset{"}{j}-1,\overset{"}{j}+1,\ldots,n \\ -\text{sign}(y_{ij})\Delta u_j & k=1,\ldots,\overset{"}{k}-1,\overset{"}{k}+1,\ldots,n \quad j=\overset{"}{j} \end{cases} \tag{I-28}$$

where the index $\overset{"}{j}$ is a different element of the set $\{1,\ldots,J-1,J+1,\ldots,n\}$ for each $k\neq\overset{"}{k}$. Now inserting (I-28) into (I-26),

$$\gamma_i \leq (1+n)\sum_{j=1}^{n}|y_{ij}|\Delta u_j \tag{I-29}$$

$$+\sum_{\substack{k=1\\k\neq\overset{"}{k}}}^{n}\left(1+\frac{n}{2}\right)\left[-|y_{i\overset{"}{j}}|\Delta u_{\overset{"}{j}} + \sum_{\substack{j=1\\j\neq\overset{"}{j}}}^{n}|y_{ij}|\Delta u_j\right]$$

$$= (1+n)\left[|y_{iJ}|\Delta u_J + \sum_{\substack{j=1\\j\neq J}}^{n}|y_{ij}|\Delta u_j\right]$$

$$+\left(1+\frac{n}{2}\right)\left[(n-1)|y_{iJ}|\Delta u_J + (n-3)\sum_{\substack{j=1\\j\neq J}}^{n}|y_{ij}|\Delta u_j\right]$$

$$= \left(\frac{n^2}{2}+\frac{3n}{2}\right)|y_{iJ}|\Delta u_J + \left(\frac{n^2}{2}+\frac{n}{2}-2\right)\sum_{\substack{j=1\\j\neq J}}^{n}|y_{iJ}|\Delta u_j$$

$$= (n+2)|y_{iJ}|\Delta u_J + \left(\frac{n^2}{2}+\frac{n}{2}-2\right)\sum_{j=1}^{n}|y_{ij}|\Delta u_j.$$

The above expression is an upper bound on the width of the tracking cycle.

Choice of the ϵ_i

Now let the ϵ_i of (5-37) be given by the γ_i upper bound:

$$\epsilon_i = (n+2)|y_{iJ}|\Delta u_J + \left(\frac{n^2}{2} + \frac{n}{2} - 2\right) \sum_{j=1}^{n} |y_{ij}|\Delta u_j. \qquad \text{(I-30)}$$

This choice, given in Chapter V as (5-38), insures satisfaction of the two conditions given below (5-37). For suppose $\underline{u}_{a-1}$ does not satisfy (5-37),

$$\underline{y}_i^T \underline{u}_{a-1} - b_i > \epsilon_i, \qquad \text{(I-31)}$$

and as a consequence the ith constraint is ignored in the computation of $\underline{u}_a$. Then $\underline{u}_a$ will not violate the ith constraint, for, defining the vector $\underline{c}_i''$ by

$$c_{ij}'' = \text{sign}(y_{ij})\Delta u_j \qquad j=1,\ldots,n, \qquad \text{(I-32)}$$

the ith constraint evaluated at $\underline{u}_a$ is

$$\underline{y}_i^T \underline{u}_a - b_i \geq \underline{y}_i^T(\underline{u}_{a-1} - \underline{c}_i'') - b_i \qquad \text{(I-33)}$$

$$= \underline{y}_i^T \underline{u}_{a-1} - b_i - \sum_{j=1}^{n} |y_{ij}|\Delta u_j$$

$$\geq \underline{y}_i^T \underline{u}_{a-1} - b_i - \sum_{j=1}^{n} |y_{ij}| \Delta u_j$$

$$- (n+2)|y_{iJ}|\Delta u_J - \left(\frac{n^2}{2} + \frac{n}{2} - 3\right) \sum_{j=1}^{n} |y_{ij}| \Delta u_j$$

$$= \underline{y}_i^T \underline{u}_{a-1} - b_i - (n+2)|y_{iJ}|\Delta u_J - \left(\frac{n^2}{2} + \frac{n}{2} - 2\right) \sum_{j=1}^{n} |y_{ij}| \Delta u_j$$

$$= \underline{y}_i^T \underline{u}_{a-1} - b_i - \epsilon_i > 0.$$

And alternatively, suppose $\underline{u}_{a-1}$ satisfies (5-37) for i=1,...,q. Then all of the $\underline{u}_{(a-1)+\beta}$, $\beta=0,\ldots,\beta^*-1$, satisfy (5-37) for i=1,...,q. For with the assumption (H-19), (5-37) evaluated at $\underline{u}_{(a-1)+\beta}$ is

$$\underline{y}_i^T \underline{u}_{(a-1)+\beta} - b_i \leq E''_{\beta a_i} + \gamma_i = \gamma_i \leq \epsilon_i. \qquad \text{(I-34)}$$

Thus, the two conditions given below (5-37) are satisfied.

APPENDIX J

DERIVATION OF MEASUREMENT NOISE EQUATIONS

The derivation of the mean and standard deviation of the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$ utilizes a number of results from a text by Papoulis.[46] References to equations from that text are denoted by the prefix "P."

Mean of Per-Cycle Increase

The first step in establishing the mean of the per-cycle increase is to insert (4-14) into (6-1) to obtain

$$\left.\frac{\partial \hat{\underline{H}}}{\partial u_{ja}}\right|_{a-1} = \left.\frac{\partial \underline{H}}{\partial u_{ja}}\right|_{a-1} + \frac{\underline{X}^{j-1,a-1} - \underline{X}^{j,a-1}}{u_{j,a-1} - u_{j,a-2}} . \qquad (J\text{-}1)$$

Then inserting (J-1) into (6-2),

$$\left.\frac{\partial \hat{F}}{\partial u_{ja}}\right|_{a-1} = \sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i}\left[\left.\frac{\partial H_i}{\partial u_{ja}}\right|_{a-1} + \frac{X_i^{j-1,a-1} - X_i^{j,a-1}}{u_{j,a-1} - u_{j,a-2}}\right] + \frac{\partial G}{\partial u_{ja}} \qquad (J\text{-}2)$$

$$= \sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i}\left.\frac{\partial H_i}{\partial u_{ja}}\right|_{a-1} + \sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i}\frac{X_i^{j-1,a-1} - X_i^{j,a-1}}{u_{j,a-1} - u_{j,a-2}} + \frac{\partial G}{\partial u_{ja}} .$$

To simplify notation, let

$$D_{ja} = \left.\frac{\partial \hat{F}}{\partial u_{ja}}\right|_{a-1}. \tag{J-3}$$

Then inserting (J-2) and (J-3) into (6-5) and taking the expected value,

$$E\{\Delta F\} = \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j E\{\text{sign}(D_{ja})\}. \tag{J-4}$$

Then using (P5-24),

$$E\{\text{sign}(D_{ja})\} = (+1)P\{\text{sign}(D_{ja})=+1\} + (-1)P\{\text{sign}(D_{ja})=-1\} \tag{J-5}$$

$$= P\{\text{sign}(D_{ja})=+1\} - (1-P\{\text{sign}(D_{ja})=+1\})$$

$$= 2P\{\text{sign}(D_{ja})=+1\} - 1$$

$$= 2P\{D_{ja}\geq 0\} - 1.$$

And using (P5-33),

$$P\{D_{ja}\geq 0\} = P\{D_{ja}\geq 0 \,|\, e_{j,a-1}=+1\}P\{e_{j,a-1}=+1\} \tag{J-6}$$

$$+ P\{D_{ja}\geq 0 \,|\, e_{j,a-1}=-1\}P\{e_{j,a-1}=-1\}.$$

Since the dependence of D_{ja} on $e_{j,a-1}$ is through the relation

$$u_{j,a-1} - u_{j,a-2} = e_{j,a-1}\Delta u_j, \tag{J-7}$$

and $e_{j,a-1}$ depends statistically only on $\chi_i^{j-1,a-2}$ and $\chi_i^{j,a-2}$, and the only coupling of measurement noise between the (a-1)st and ath algorithmic cycles is through the relation

$$\underline{\chi}^{n,a-1} = \underline{\chi}^{0,a-2}, \tag{J-8}$$

it follows that D_{ja} and $e_{j,a-1}$ are statistically related only if n=1. Thus, for n≥2, (J-6) may be written

$$P\{D_{ja}\geq 0\} = P\{D_{ja}^{+}\geq 0\}P\{e_{j,a-1}=+1\} + P\{D_{ja}^{-}\geq 0\}P\{e_{j,a-1}=-1\} \tag{J-9}$$

where D_{ja}^{+} and D_{ja}^{-} are, respectively, D_{ja} with $(u_{j,a-1}-u_{j,a-2})$ replaced by $+\Delta u_j$ and $-\Delta u_j$.

Since D_{ja}^{+} and D_{ja}^{-} are linear combinations of Gaussian random variables, they themselves are Gaussian random variables. Their means are

$$E\{D_{ja}^{\pm}\} = \sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i} \left.\frac{\partial H_i}{\partial u_{ja}}\right|_{a-1} + \frac{\partial G}{\partial u_{ja}} \tag{J-10}$$

and their variances are

$$E\left\{(D_{ja}^{\pm}-E\{D_{ja}^{\pm}\})^2\right\} = E\left\{\left(\sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i} \frac{\chi_i^{j-1,a-1}-\chi_i^{j,a-1}}{\pm \Delta u_j}\right)^2\right\} \tag{J-11}$$

$$= \frac{1}{(\Delta u_j)^2} E\left\{\left(\sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i} (\chi_i^{j-1,a-1}-\chi_i^{j,a-1})\right)^2\right\}$$

$$= \frac{1}{(\Delta u_j)^2} E\left\{ \sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 (\chi_i^{j-1,a-1} - \chi_i^{j,a-1})^2 \right.$$

$$\left. + \sum_{i=1}^{\mu} \sum_{\substack{k=1 \\ k\neq i}}^{\mu} \frac{\partial G}{\partial H_i} \frac{\partial G}{\partial H_k} (\chi_i^{j-1,a-1} - \chi_i^{j,a-1})(\chi_k^{j-1,a-1} - \chi_k^{j,a-1}) \right\}$$

$$= \frac{1}{(\Delta u_j)^2} E\left\{ \sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 (\chi_i^{j-1,a-1} - \chi_i^{j,a-1})^2 \right\}$$

$$= \frac{1}{(\Delta u_j)^2} \sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 2\sigma_i^2.$$

Since D_{ja}^{+} and D_{ja}^{-} have the same mean and variance,

$$P\{D_{ja}^{+} \geq 0\} = P\{D_{ja}^{-} \geq 0\}. \qquad \text{(J-12)}$$

Thus, (J-9) may be written

$$P\{D_{ja} \geq 0\} = P\{D_{ja}^{+} \geq 0\}(P\{e_{j,a-1} = +1\} + P\{e_{j,a-1} = -1\}) \qquad \text{(J-13)}$$

$$= P\{D_{ja}^{+} \geq 0\}.$$

Inserting (P3-32) and (P4-41) into (J-13),

$$P\{D_{ja}^{+} \geq 0\} = \frac{1}{2} + \text{erf}\left(\frac{E\{D_{ja}^{+}\}}{\sqrt{E\{(D_{ja}^{+} - E\{D_{ja}^{+}\})^2\}}}\right) \qquad \text{(J-14)}$$

where the function erf(x) is given by (P3-31) as

$$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2}\, dy. \qquad \text{(J-15)}$$

Finally, combining (J-4), (J-5), (J-13), and (J-14),

$$E\{\Delta F\} = \sum_{j=1}^{n} \left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1} \Delta u_j 2\text{erf}\left[\frac{\left(\sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i} \left.\frac{\partial H_i}{\partial u_{ja}}\right|_{a-1} + \frac{\partial G}{\partial u_{ja}}\right)\Delta u_j}{\sqrt{\sum_{i=1}^{\mu}\left(\frac{\partial G}{\partial H_i}\sqrt{2}\,\sigma_i\right)^2}}\right]. \qquad \text{(J-16)}$$

This is the equation given in Chapter VI as (6-6).

Standard Deviation of Per-Cycle Increase

Before considering the standard deviation of the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$, consider the variance, denoted by $\text{Var}\{\Delta F\}$. From (P5-36),

$$\text{Var}\{\Delta F\} = E\{(\Delta F)^2\} - (E\{\Delta F\})^2. \qquad \text{(J-17)}$$

Denoting the argument of erf(·) in (J-16) by A_j, it follows from (J-16) that

$$(E\{\Delta F\})^2 = \left(\sum_{j=1}^{n} \left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \Delta u_j 2\mathrm{erf}(A_j) \right)^2 \tag{J-18}$$

$$= \sum_{j=1}^{n} \left(\left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \Delta u_j \right)^2 4\mathrm{erf}^2(A_j)$$

$$+ \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq j}}^{n} \left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \left. \frac{\partial F}{\partial u_{ka}} \right|_{a-1} \Delta u_j \Delta u_k 4\mathrm{erf}(A_j)\mathrm{erf}(A_k).$$

And from (J-4),

$$E\{(\Delta F)^2\} = E\left\{ \left(\sum_{j=1}^{n} \left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \Delta u_j \mathrm{sign}(D_{ja}) \right)^2 \right\} \tag{J-19}$$

$$= E\left\{ \sum_{j=1}^{n} \left(\left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \Delta u_j \right)^2 \right.$$

$$\left. + \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq j}}^{n} \left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \left. \frac{\partial F}{\partial u_{ka}} \right|_{a-1} \Delta u_j \Delta u_k \mathrm{sign}(D_{ja})\mathrm{sign}(D_{ka}) \right\}$$

$$= \sum_{j=1}^{n} \left(\left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \Delta u_j \right)^2$$

$$+ \sum_{j=1}^{n} \sum_{\substack{k=1 \\ k \neq j}}^{n} \left. \frac{\partial F}{\partial u_{ja}} \right|_{a-1} \left. \frac{\partial F}{\partial u_{ka}} \right|_{a-1} \Delta u_j \Delta u_k E\{\mathrm{sign}(D_{ja})\mathrm{sign}(D_{ka})\}.$$

Now, as with the computation of the mean,

$$E\{\text{sign}(D_{ja})\text{sign}(D_{ka})\} \tag{J-20}$$

$$= (+1)P\{\text{sign}(D_{ja})\text{sign}(D_{ka})=+1\} + (-1)P\{\text{sign}(D_{ja})\text{sign}(D_{ka})=-1\}$$

$$= P\{\text{sign}(D_{ja})\text{sign}(D_{ka})=+1\} - (1-P\{\text{sign}(D_{ja})\text{sign}(D_{ka})=+1\})$$

$$= 2P\{\text{sign}(D_{ja})\text{sign}(D_{ka})=+1\} - 1$$

$$= 2P\{D_{ja}D_{ka}\geq 0\} - 1$$

and

$$P\{D_{ja}D_{ka}\geq 0\} \tag{J-21}$$

$$= P\{D_{ja}D_{ka}\geq 0 \,|\, e_{j,a-1}=+1 \text{ and } e_{k,a-1}=+1\}P\{e_{j,a-1}=+1 \text{ and } e_{k,a-1}=+1\}$$

$$+ P\{D_{ja}D_{ka}\geq 0 \,|\, e_{j,a-1}=-1 \text{ and } e_{k,a-1}=-1\}P\{e_{j,a-1}=-1 \text{ and } e_{k,a-1}=-1\}$$

$$+ P\{D_{ja}D_{ka}\geq 0 \,|\, e_{j,a-1}=+1 \text{ and } e_{k,a-1}=-1\}P\{e_{j,a-1}=+1 \text{ and } e_{k,a-1}=-1\}$$

$$+ P\{D_{ja}D_{ka}\geq 0 \,|\, e_{j,a-1}=-1 \text{ and } e_{k,a-1}=+1\}P\{e_{j,a-1}=-1 \text{ and } e_{k,a-1}=+1\}.$$

For $n\geq 2$, (J-21) may be simplified to

$$P\{D_{ja}D_{ka}\geq 0\} = P\{D_{ja}^{+}D_{ka}^{+}\geq 0\}P\{e_{j,a-1}=+1 \text{ and } e_{k,a-1}=+1\} \qquad \text{(J-22)}$$

$$+ P\{D_{ja}^{-}D_{ka}^{-}\geq 0\}P\{e_{j,a-1}=-1 \text{ and } e_{k,a-1}=-1\}$$

$$+ P\{D_{ja}^{+}D_{ka}^{-}\geq 0\}P\{e_{j,a-1}=+1 \text{ and } e_{k,a-1}=-1\}$$

$$+ P\{D_{ja}^{-}D_{ka}^{+}\geq 0\}P\{e_{j,a-1}=-1 \text{ and } e_{k,a-1}=+1\}.$$

For $k=1,\ldots,j-2,j+2,\ldots,n$, $D_{ja}^{\pm}$ and $D_{ka}^{\pm}$ are independent of one another. Hence

$$P\{D_{ja}^{\pm}D_{ka}^{\pm}\geq 0\} = P\{D_{ja}^{\pm}\geq 0\}P\{D_{ka}^{\pm}\geq 0\} + P\{D_{ja}^{\pm}\leq 0\}P\{D_{ka}^{\pm}\leq 0\}. \qquad \text{(J-23)}$$

Since D_{ja}^{+} and D_{ja}^{-} have the same mean and variance,

$$P\{D_{ja}^{+}\gtrless 0\} = P\{D_{ja}^{-}\gtrless 0\}$$

$$P\{D_{ka}^{+}\gtrless 0\} = P\{D_{ka}^{-}\gtrless 0\}. \qquad \text{(J-24)}$$

Thus, (J-22) may be simplified to

$$P\{D_{ja}D_{ka}\geq 0\} = P\{D_{ja}^{+}\geq 0\}P\{D_{ka}^{+}\geq 0\} + P\{D_{ja}^{+}\leq 0\}P\{D_{ka}^{+}\leq 0\} \qquad \text{(J-25)}$$

$$= \left(\frac{1}{2}+\text{erf}(A_j)\right)\left(\frac{1}{2}+\text{erf}(A_k)\right) + \left(\frac{1}{2}-\text{erf}(A_j)\right)\left(\frac{1}{2}-\text{erf}(A_k)\right)$$

$$= \frac{1}{2} + 2\text{erf}(A_j)\text{erf}(A_k) \qquad k=1,\ldots,j-2,j+2,\ldots,n.$$

But for k=j-1,j+1, $D_{ja}^{\pm}$ and $D_{ka}^{\pm}$ are not independent of one another and the probability $P\{D_{ja}^{\pm} D_{ka}^{\pm} \geq 0\}$ is governed by a joint density function. The marginal statistics of that function (see P6-65) are given by (J-10) and (J-11). The correlation coefficient, from (P7-66), is

$$r^{\pm\pm} = \frac{E\left\{\sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i} \frac{\chi_i^{j-1,a-1} - \chi_i^{j,a-1}}{\pm\Delta u_j} \sum_{i=1}^{\mu} \frac{\partial G}{\partial H_i} \frac{\chi_i^{k-1,a-1} - \chi_i^{k,a-1}}{\pm\Delta u_k}\right\}}{\sqrt{\frac{1}{(\Delta u_j)^2} \sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 2\sigma_i^2} \sqrt{\frac{1}{(\Delta u_k)^2} \sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 2\sigma_i^2}} \qquad \text{(J-26)}$$

$$= \frac{(\pm 1)(\pm 1)}{\sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 2\sigma_i^2} E\left\{\sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 (\chi_i^{j-1,a-1} - \chi_i^{j,a-1})(\chi_i^{k-1,a-1} - \chi_i^{k,a-1}) \right.$$

$$\left. + \sum_{i=1}^{\mu} \sum_{\substack{i=1 \\ i \neq i}}^{\mu} \frac{\partial G}{\partial H_i} \frac{\partial G}{\partial H_i} (\chi_i^{j-1,a-1} - \chi_i^{j,a-1})(\chi_i^{k-1,a-1} - \chi_i^{k,a-1})\right\}$$

$$= \frac{(\pm 1)(\pm 1)}{\sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 2\sigma_i^2} E\left\{\sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 (\chi_i^{j-1,a-1} - \chi_i^{j,a-1})(\chi_i^{k-1,a-1} - \chi_i^{k,a-1})\right\}$$

$$= \frac{(\pm 1)(\pm 1)}{\sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 2\sigma_i^2} \sum_{i=1}^{\mu} \left(\frac{\partial G}{\partial H_i}\right)^2 (-\sigma_i^2) = \frac{-(\pm 1)(\pm 1)}{2} .$$

Since the marginal statistics of $P\{D^{\pm}_{ja} D^{\pm}_{ka} \geq 0\}$ are the same no matter what the ± combination,

$$P\{D^{+}_{ja} D^{+}_{ka} \geq 0\} = P\{D^{-}_{ja} D^{-}_{ka} \geq 0\}$$
$$P\{D^{+}_{ja} D^{-}_{ka} \geq 0\} = P\{D^{-}_{ja} D^{+}_{ka} \geq 0\}. \tag{J-27}$$

It follows that (J-22) may be simplified to

$$\begin{aligned} P\{D_{ja} D_{ka} \geq 0\} &= P\{D^{+}_{ja} D^{+}_{ka} \geq 0\} P\{e_{j,a-1} e_{k,a-1} = +1\} \\ &\quad + P\{D^{+}_{ja} D^{-}_{ka} \geq 0\} P\{e_{j,a-1} e_{k,a-1} = -1\} \\ &= P\{D^{+}_{ja} D^{+}_{ka} \geq 0\} P\{e_{j,a-1} e_{k,a-1} = +1\} \\ &\quad + P\{D^{+}_{ja} + D^{-}_{ka} \geq 0\} (1 - P\{e_{j,a-1} e_{k,a-1} = +1\}) \\ &= P\{D^{+}_{ja} D^{-}_{ka} \geq 0\} \\ &\quad + P\{e_{j,a-1} e_{k,a-1} = +1\} (P\{D^{+}_{ja} D^{+}_{ka} \geq 0\} - P\{D^{+}_{ja} D^{-}_{ka} \geq 0\}). \end{aligned} \tag{J-28}$$

But since

$$P\{e_{j,a-1} e_{k,a-1} = +1\} = P\{D_{j,a-1} D_{k,a-1} \geq 0\}, \tag{J-29}$$

(J-28) may be written

$$P\{D_{ja}D_{ka}\geq 0\} = P\{D_{ja}^{+}D_{ka}^{-}\geq 0\} \tag{J-30}$$

$$+ P\{D_{j,a-1}D_{k,a-1}\geq 0\}(P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} - P\{D_{ja}^{+}D_{ka}^{-}\geq 0\}).$$

If the joint statistics of $D_{ja}^{\pm}$ and $D_{ka}^{\pm}$ are time-invariant, (J-30) provides a recursive relation for evaluating $P\{D_{ja}D_{ka}\geq 0\}$. The result of recursively inserting $P\{D_{j,a-1}D_{k,a-1}\geq 0\}$, $P\{D_{j,a-2}D_{k,a-2}\geq 0\},\ldots,$ and $P\{D_{j1}D_{k1}\geq 0\}$ into (J-30) is

$$P\{D_{ja}D_{ka}\geq 0\} = \sum_{\alpha=0}^{a-2} P\{D_{ja}^{+}D_{ka}^{-}\geq 0\}(P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} - P\{D_{ja}^{+}D_{ka}^{-}\geq 0\})^{\alpha} \tag{J-31}$$

$$+ P\{D_{j1}D_{k1}\geq 0\}(P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} - P\{D_{ja}^{+}D_{ka}^{+}\geq 0\})^{a-1}.$$

In the limit as $a \to \infty$, the second term above approaches zero since

$$\left|P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} - P\{D_{ja}^{+}D_{ka}^{-}\geq 0\}\right| < 1, \tag{J-32}$$

and the first term approaches the sum of an infinite geometric series. Thus, for $a \to \infty$,

$$P\{D_{ja}D_{ka}\geq 0\} = \frac{P\{D_{ja}^{+}D_{ka}^{-}\geq 0\}}{1 - (P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} - P\{D_{ja}^{+}D_{ka}^{-}\geq 0\})} \tag{J-33}$$

$$k=j-1,j+1.$$

Finally, combining (J-17)-(J-20), (J-25), and (J-33),

$$\mathrm{Var}\{\Delta F\} = \sum_{j=1}^{n}\left(\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\Delta u_j\right)^2\left(1-4\mathrm{erf}^2(A_j)\right) \qquad \text{(J-34)}$$

$$+ \sum_{j=1}^{n}\sum_{\substack{k=1\\k=j-1,j+1}}^{n}\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\left.\frac{\partial F}{\partial u_{ka}}\right|_{a-1}\Delta u_j\Delta u_k \times$$

$$\left[\frac{P\{D_{ja}^{+}D_{ka}^{-}\geq 0\} + P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} - 1}{P\{D_{ja}^{+}D_{ka}^{-}\geq 0\} - P\{D_{ja}^{+}D_{ka}^{+}\geq 0\} + 1} - 4\mathrm{erf}(A_j)\mathrm{erf}(A_k)\right]$$

$$= \sum_{j=1}^{n}\left(\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\Delta u_j\right)^2\left(1-4\mathrm{erf}^2(A_j)\right)$$

$$+ 2\sum_{j=1}^{n-1}\left.\frac{\partial F}{\partial u_{ja}}\right|_{a-1}\left.\frac{\partial F}{\partial u_{j+1,a}}\right|_{a-1}\Delta u_j\Delta u_{j+1} \times$$

$$\left[\frac{P\{D_{ja}^{+}D_{j+1,a}^{-}\geq 0\} + P\{D_{ja}^{+}D_{j+1,a}^{+}\geq 0\} - 1}{P\{D_{ja}^{+}D_{j+1,a}^{-}\geq 0\} - P\{D_{ja}^{+}D_{j+1,a}^{+}\geq 0\} + 1} - 4\mathrm{erf}(A_j)\mathrm{erf}(A_{j+1})\right].$$

The standard deviation of the per-cycle increase in $F(u_{1a},\ldots,u_{na},t^a)$, denoted by $SD\{\Delta F\}$, is the positive square root of the above expression. The results of a numerical evaluation of this expression are given in Figure 18.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. M. D. Mesarovic, "Multilevel Systems and Concepts in Process Control, *IEEE Proceedings*, vol. 58, pp. 111-125, 1970.

2. H. E. Pike, "Process Control Software," *IEEE Proceedings*, vol. 58, pp. 87-97, 1970.

3. M. G. Slin'ko, Y. M. Buzhdan, V. S. Beskov, and I. D. Emel'yanov, "Optimum Conditions for Preparing Ethylene Oxide," *Kinetics and Catalysis*, vol. 3, pp. 121-128, 1962.

4. C. H. Tsai, *Optimum Constrained Control of an Unsteady-State Tubular Reactor*, Ph.D. Thesis, Iowa State University of Science and Technology, 1967.

5. O. Bilous and N. R. Amundson, "Optimum Temperature Gradients in Tubular Reactors I. General Theory and Methods," *Chemical Engineering Science*, vol. 5, pp. 81-92, 1956.

6. O. Bilous and N. R. Amundson, "Optimum Temperature Gradients in Tubular Reactors II. Numerical Study," *Chemical Engineering Science*, vol. 5, pp. 115-126, 1956.

7. R. E. Gee, W. H. Linton, R. E. Maier, and J. W. Raines, "Use of Computers in Kinetic Calculations, Gas-Phase Tubular Reactor Kinetics Involving Differential Fouling of Heat Transfer Surface," *Chemical Engineering Progress*, vol. 50, pp. 497-502, 1954.

8. M. R. Newberger, *Optimal Operation of a Tubular Reactor*, Ph.D. Thesis, University of Michigan, 1968.

9. M. D. Rafal and J. S. Dranoff, "Efficient Algorithm for Optimization of a Multibed Adiabatic Reactor Sequence," *Industrial and Engineering Chemistry--Process Design and Development*, vol. 5, pp. 129-135, 1966.

10. L. M. Pis'men and I. I. Ioffe, "Calculation of the Optimal Process in Chemical Reactors by the Method of Dynamic Programming," *International Chemical Engineering*, vol. 3, pp. 24-32, 1963.

11. R. P. King, "Calculation of the Optimal Conditions for Chemical Reactors of the Combined Type," *Chemical Engineering Science*, vol. 20, pp. 537-544, 1965.

12. E. D. Crandall and W. F. Stevens, "An Application of Adaptive Control to a Continuous Stirred-Tank Reactor," *Journal of the American Institute of Chemical Engineers*, vol. 11, pp. 930-936, 1965.

13. A. Chou, W. H. Ray, and R. Aris, "Simple Control Policies for Reactors with Catalyst Decay," *Trans. Institution of Chemical Engineers*, vol. 45, pp. T153-T159, 1967.

14. G. R. Gavalas and J. H. Seinfeld, "Sequential Estimation of States and Kinetic Parameters in Tubular Reactors with Catalyst Decay," *Chemical Engineering Science*, vol. 24, pp. 625-636, 1969.

15. D. R. Bertran and K. S. Chang, "Optimal Feedforward Control of Concurrent Tubular Reactors," *Journal of the American Institute of Chemical Engineers*, vol. 16, pp. 897-902, 1970.

16. C. H. Wells and R. E. Larson. "Application of Combined Optimum Control and Estimation Theory to Direct Digital Control," *IEEE Proceedings*, vol. 58, pp. 16-22, 1970.

17. H. Witsenhausen, "Development of a Program for the Hybrid Simulation of a Tubular Reactor," *Proceedings of the International Association for Analog Computation*, vol. 6, pp. 112-117, 1964.

18. A. E. Foureau, "A Study of the Automatic Control of a Tubular Reactor," *Proceedings of the International Federation of Automatic Control*, paper 38b, London, 1966.

19. R. N. Schindler and R. Aris, "Questing Control of a Stirred Tank Reactor," *Chemical Engineering Science*, vol. 22, pp. 319-336, 1967.

20. R. N. Schindler and R. Aris, "The Questing Control of a Two Phase Reactor," *Chemical Engineering Science*, vol. 22, pp. 337-344, 1967.

21. R. N. Schindler and R. Aris, "Questing Control with an Economic Criterion," *Chemical Engineering Science*, vol. 22, pp. 345-352, 1967.

22. V. W. Eveleigh, *Adaptive Control and Optimization Techniques*, McGraw-Hill Book Company, New York, 1967.

23. M. H. Hamza, "Extremum Control in the Presence of Variable Pure Delay," *Proc. IFAC Symposium on Theory of Self-Adaptive Control Systems*, pp. 290-298, Teddington, 1965.

24. C. M. Woodside, "The Design of Optimal Extremum Controllers," *International Journal of Control*, vol. 8, pp. 545-559, 1968.

25. S. M. Roberts and H. I. Lyvers, "The Gradient Method in Process Control," *Industrial and Engineering Chemistry*, vol. 53, pp. 877-882, 1961.

26. B. Kondo and T. Suzuki, "Dynamics of Optimizing Control Systems Based on the Alternate-Parameter Adjustment Method," *Electrical Engineering Japan*, vol. 84, no. 5, pp. 64-72, 1964.

27. A. Gersho, "Adaptation in a Quantized Parameter Space," *Proc. 6th Allerton Conference on Circuit Theory and System Theory*, pp. 646-653, Monticello, Ill., 1968.

28. O. L. R. Jacobs and W. M. Wonham, "Extremum Control in the Presence of Noise," *Journal of Electronics and Control*, vol. 11, pp. 193-211, 1961.

29. J. J. Florentine, "An Approximately Optimal Extremum Regulator," *Journal of Electronics and Control*, vol. 17, pp. 211-222, 1964.

30. M. Orban, "Some Questions of Learning Optimization of Large-Scale Processes," *Proc. IFAC Symposium on Theory of Self-Adaptive Control Systems*, pp. 309-321, Teddington, 1965.

31. G. Zoutendijk, "Nonlinear Programming: A Numerical Survey," *Journal SIAM Control*, vol. 4, pp. 194-210, 1966.

32. J. V. Flynn and L. Lapidus, "The Control of Nonlinear Systems VI. Computational Use of Nonlinear Programming," *Journal of the American Institute of Chemical Engineers*, vol. 15, pp. 308-311, 1969.

33. N. R. Amundson and D. Luss, "Qualitative and Quantitative Observations on the Tubular Reactor," *Canadian Journal of Chemical Engineering*, vol. 46, pp. 424-443, 1968.

34. R. Aris, *Introduction to the Analysis of Chemical Reactors*, Prentice-Hall, Englewood Cliffs, N. J., 1965.

35. O. Levenspiel, *Chemical Reaction Engineering*, John Wiley & Sons, New York, 1962.

36. J. S. Berger and D. D. Perlmutter, "Chemical Reactor Stability by Liapunov's Direct Method," *Journal of the American Institute of Chemical Engineers*, vol. 10, pp. 233-238, 1964.

37. J. S. Berger and D. D. Perlmutter, "The Effect of Feedback Control on Chemical Reactor Stability," *Journal of the American Institute of Chemical Engineers*, vol. 10, pp. 238-245, 1964.

38. S. Katz, "Best Temperature Profiles in Plug-Flow Reactions: Method of the Calculus of Variations," *Annals of the New York Academy of Science*, vol. 84, pp. 441-478, 1960.

39. R. D. McCoy and B. O. Ayers, "On-Line Stream Analysis with a Chromatograph," *Control Engineering*, vol. 17, no. 7, pp. 44-49, 1970.

40. D. E. Seborg and E. F. Johnson, "Construction of Regions of Stability for Nonlinear Systems Containing Time Delays," *Preprints of the Joint Automatic Control Conference*, paper 10-B, Atlanta, 1970.

41. R. Vichnevetsky, "Method of Characteristics in the Hybrid Solution of First Order Partial Differential Equations," *ECC Report No. 60*, Electronics Associates, Inc., 1962.

42. E. A. Coddington, *An Introduction to Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1961.

43. J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part I. Linear Constraints," *Journal SIAM*, vol. 8, pp. 181-217, 1960.

44. J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming, Part II. Nonlinear Constraints," *Journal SIAM*, vol. 9, pp. 514-532, 1961.

45. W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1964.

46. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

47. D. J. Wilde and C. S. Beightler, *Foundations of Optimization*, Prentice-Hall, Englewood Cliffs, N. J., 1967.

48. P. Harriott, *Process Control*, McGraw-Hill, New York, 1964.

49. P. Fangel, "A General Method of Dimensioning the Temperature Control System of a Continuous-Flow Stirred-Tank Reactor," *Chemical Engineering Science*, vol. 21, pp. 49-61, 1966.

50. F. A. Fine and S. C. Bankoff, "Control Vector Iteration in Chemical Plant Optimization," *Industrial and Engineering Chemistry--Fundamentals*, vol. 6, pp. 288-293, 1967.

51. B. F. Rothenberger and L. Lapidus, "The Control of Nonlinear Systems. Quasilinearization and State-Constrained Systems," *Journal of the American Institute of Chemical Engineers*, vol. 13, pp. 982-988, 1967.

52. W. O. Paradis and D. D. Perlmutter, "Optimality and Computational Feasibility in Transient Control II. Feedback Control of a Distributed Parameter Process," *Journal of the American Institute of Chemical Engineers*, vol. 12, pp. 883-890, 1966.

53. R. F. Barrett, C. O. Alford, and J. L. Hammond, "On-Line Optimization of the Multi-Jacketed Plug-Flow Tubular Chemical Reactor," *Proc. 2nd S.E. Symposium on System Theory*, paper B2, Gainesville, Florida, 1970.

54. D. Luss and N. R. Amundson, "Some General Observations on Tubular Reactor Stability," *Canadian Journal of Chemical Engineering*, vol. 45, pp. 341-346, 1967.

55. D. Luss, "Sufficient Conditions for Uniqueness of the Steady-State Solutions in Distributed Parameter Systems," *Chemical Engineering Science*, vol. 23, pp. 1249-1255, 1968.

56. A. J. Berger and L. Lapidus, "An Introduction to the Stability of Distributed Systems via a Liapunov Functional," *Journal of the American Institute of Chemical Engineers*, vol. 14, pp. 558-568, 1968.

57. L. T. Fan and R. C. Bailie, "Axial Diffusion in Isothermal Tubular Flow Reactors," *Chemical Engineering Science*, vol. 13, pp. 63-68, 1960.

58. S. L. Liu, "Numerical Solution of Two-Point Boundary Value Problems in Simultaneous, Second-Order Nonlinear Differential Equations," *Chemical Engineering Science*, vol. 22, pp. 871-881, 1967.

59. H. W. Pakes and C. Storey, "Solution of the Equations for a Tubular Reactor with Axial Diffusion by a Variational Technique," *The Chemical Engineer*, no. 208, pp. CE96-CE98, CE108, 1967.

60. T. Takamatsu, I. Hashimoto, and Y. Sawanoi, "The Effect of Fluid Mixing on the Maximum Yield and the Optimum Temperature Profile in a Tubular Reactor," *Memoirs of the Faculty of Engineering, Kyoto University*, vol. 29, pp. 225-239, 1967.

61. L. Markus and N. R. Amundson, "Nonlinear Boundary-Value Problems Arising in Chemical Reactor Theory," *Journal of Differential Equations*, vol. 4, pp. 102-113, 1968.

62. J. D. Tinkler and D. E. Lamb. "Dynamics and Control of a Fixed-Bed Chemical Reactor," *Chemical Engineering Progress Symposium Series*, no. 55, pp. 155-167, 1965.

63. I. S. Sokolnikoff, "Approximate Methods of Solution of Two-Dimensional Problems in Anisotropic Elasticity," *Proceedings of Symposia in Applied Mathematics*, vol. 3, pp. 1-11, 1950.

64. J. L. Nowinski and I. A. Ismail, "Application of a Multi-Parameter Perturbation Method to Elastostatics," in *Developments in Theoretical and Applied Mechanics*, ed. W. A. Shaw, Pergamon Press, Oxford, 1965.

65. R. Bellman, *Perturbation Techniques in Mathematics, Physics, and Engineering*, Holt, Rinehart, and Winston, New York, 1964.

66. K. L. King and S. H. Davis, "The Effect of Dispersion on Chemical Reactor Optimum Residence Times Using Perturbation Expansion Techniques," *Chemical Engineering Science*, vol. 23, pp. 833-840, 1968.

67. S. K. T. Hsu and R. M. Howe, "Preliminary Investigation of a Hybrid Method for Solving Partial Differential Equations," *AFIPS Proceedings*, Fall Joint Computer Conference, pp. 601-609, 1968.

VITA

Richard Fox Barrett was born in Atlanta, Georgia, on May 8, 1943. He is the son of George Dickey Barrett and Carolyn Burns Barrett.

After graduating with honors from the Henry Grady High School in Atlanta, Georgia, in June, 1961, he attended the Georgia Institute of Technology where he received the B.E.E. degree, with highest honor, in June, 1965, and the M.S.E.E. degree in June, 1969.

From September, 1965, to September, 1968, he held a National Defense Education Act Graduate Fellowship at the Georgia Institute of Technology. He has held summer positions with the American Telephone and Telegraph Company, Conyers, Georgia, and with the E. I. du Pont de Nemours & Company, Orange, Texas.

Mr. Barrett is a member of Tau Beta Pi, Phi Kappa Phi, Eta Kappa Nu, and Sigma Xi.